

Bayesian Fit of Exclusive $b \rightarrow s \bar{\ell} \ell$ Decays: The Standard Model Operator Basis

Frederik Beaujean^a Christoph Bobeth^b Danny van Dyk^c Christian Wacker^c

^a*Max-Planck-Institut für Physik, 80805 München, Germany*

^b*Institute for Advanced Study & Excellence Cluster Universe, Technische Universität München, 85748 Garching, Germany*

^c*Institut für Physik, Technische Universität Dortmund, 44221 Dortmund, Germany*

E-mail: beaujean@mpp.mpg.de, bobeth@ph.tum.de,
danny.dyk@tu-dortmund.de, christian.wacker@tu-dortmund.de

ABSTRACT: We perform a model-independent fit of the short-distance couplings $\mathcal{C}_{7,9,10}$ within the Standard Model set of $b \rightarrow s \gamma$ and $b \rightarrow s \bar{\ell} \ell$ operators. Our analysis of $B \rightarrow K^* \gamma$, $B \rightarrow K^{(*)} \bar{\ell} \ell$ and $B_s \rightarrow \bar{\mu} \mu$ decays is the first to harness the full power of the Bayesian approach: all major sources of theory uncertainty explicitly enter as nuisance parameters. Exploiting the latest measurements, the fit reveals a flipped-sign solution in addition to a Standard-Model-like solution for the couplings \mathcal{C}_i . Each solution contains about half of the posterior probability, and both have nearly equal goodness of fit. The Standard Model prediction is close to the best-fit point. No New Physics contributions are necessary to describe the current data. Benefitting from the improved posterior knowledge of the nuisance parameters, we predict ranges for currently unmeasured, optimized observables in the angular distributions of $B \rightarrow K^* (\rightarrow K \pi) \bar{\ell} \ell$.

Contents

1	Introduction	2
2	$\Delta B=1$ Decays: Conventions, Observables and Experimental Input	4
2.1	$\Delta B=1$ Effective Theory	4
2.2	Observables and Experimental Input	5
2.2.1	$B \rightarrow K^* \gamma$	5
2.2.2	$B \rightarrow K \bar{\ell} \ell$	6
2.2.3	$B \rightarrow K^* (\rightarrow K \pi) \bar{\ell} \ell$	6
2.2.4	$B_s \rightarrow \bar{\mu} \mu$	10
3	Statistical Method	10
3.1	Monte Carlo Algorithm	11
3.2	Priors	12
3.3	Experimental Results	12
3.4	Uncertainties of Theory Predictions	13
3.5	Goodness of Fit and Model Comparison	14
4	Results	15
4.1	Fit Results	15
4.2	Predictions	23
5	Conclusion	27
A	Numerical Input	28
B	Nuisance Parameters	29
B.1	Common Nuisance Parameters	29
B.2	$B \rightarrow K^{(*)}$ Form Factors and f_{B_s}	30
B.3	Subleading Λ/m_b Corrections	32
C	Standard Model Predictions	33
D	Distributions	34
D.1	Amoroso Distribution	34
D.2	LogGamma Distribution	36

1 Introduction

In the course of the last decade, rare B -meson decays were discovered that are mediated at the parton level by flavor-changing neutral-current (FCNC) transitions $b \rightarrow s \bar{\ell} \ell$ and $b \rightarrow s \gamma$. They allow one to test Standard-Model (SM) predictions at the loop level and to conduct searches for indirect signals of physics beyond the SM (BSM), providing strong constraints on the corresponding fundamental parameters, especially in the quark flavor sector.

The radiative FCNC decay $B \rightarrow K^* \gamma$ was first observed by the CLEO collaboration at the Cornell Electron Storage Ring [1]. The first-generation B factory experiments BaBar [2–9] and Belle [10–13] observed rare radiative and semileptonic FCNC decays of the B meson with branching fractions of 10^{-4} to 10^{-7} . They measured branching ratios and spectral information for a set of inclusive $B \rightarrow X_s \bar{\ell} \ell$ and exclusive $B \rightarrow K^{(*)} \bar{\ell} \ell$ ($\ell = e, \mu$) decays. Recently, additional exclusive $B \rightarrow K^{(*)} \bar{\mu} \mu$ decay modes were measured by the hadron collider experiments CDF [14–16] at the Tevatron and LHCb [17, 18] at the Large Hadron Collider (LHC). The complete analyses of the full BaBar, Belle and CDF data sets is expected to be published soon. LHCb is about to significantly improve the accuracy of measurements of exclusive decays, eventually dominating the other experiments in terms of collected numbers of events by the end of 2012. The multipurpose LHC experiments ATLAS and CMS are expected to perform similar searches.

In the last decade DØ [19] and CDF [20, 21] significantly improved the upper bound on the branching ratio of the very rare leptonic decay $B_s \rightarrow \bar{\mu} \mu$ by several orders of magnitude. Currently, LHCb, CMS, and ATLAS continue this search [22–26], and a future discovery at SM rates of about 3×10^{-9} is possible with sufficient luminosity [27].

Theory predictions of the inclusive decay $B \rightarrow X_s \bar{\ell} \ell$ have reached the next-to-next-to-leading order (NNLO) [28–34]. Contrary to exclusive decays, it only depends on nonperturbative hadronic matrix elements at subleading order in the Heavy Quark Expansion. But the current measurements of its branching fraction are still very uncertain [2, 11] and provide only very limited spectral information in the dilepton invariant mass. This situation is not likely to improve until the end of the run of the superflavor factories Belle II [35] and possibly SuperB [36] around the year 2020. Therefore, it is desirable to include exclusive decays in tests of the SM and searches for BSM signals, especially in view of the high number of events expected at LHCb. Moreover, the angular distribution of the exclusive decay $B \rightarrow K^* (\rightarrow K \pi) \bar{\ell} \ell$ with a 4-body final state offers a multitude of optimized observables — see [37] for a short summary.

Both inclusive and exclusive decays are described by the $\Delta B=1$ effective theory of electroweak interactions of the SM and its extensions. They provide constraints on the effective short-distance couplings which are known precisely in the SM and are the main objects of interest due to their sensitivity to BSM effects at the electroweak scale.

Exclusive decays typically require the inclusion of final-state-specific nonperturbative (hadronic) QCD effects, which complicate the extraction of the short-distance couplings. Analyses are further complicated by the background processes $b \rightarrow s + (\bar{q} q) \rightarrow s + \bar{\ell} \ell$ induced by 4-quark operators $b \rightarrow s \bar{q} q$ ($q = u, d, s, c$). In particular, the narrow J/ψ and ψ' resonances constitute huge backgrounds to the short-distance-dominated $b \rightarrow s \bar{\ell} \ell$ processes.

Consequently, theory predictions focus on the q^2 regions below and above both resonances and are usually referred to as being in the low- or high- q^2 regions, or at large and low hadronic recoil. QCD factorization (QCDF) [38–41] or Soft Collinear Effective Theory (SCET) [42, 43] is applied at large recoil, $E \sim m_b$, of the $K^{(*)}$ system, typically in the range $1 \text{ GeV}^2 \lesssim q^2 \lesssim 6 \text{ GeV}^2$. The expansion parameter $\lambda = \Lambda/E$ is of the order Λ/m_b , resulting in a double expansion in λ and the QCD coupling constant α_s . Here Λ denotes a scale associated with nonperturbative QCD dynamics, typically $\lesssim 500 \text{ MeV}$. The exact interpretation is process dependent and specific to the expansion. An operator product expansion (OPE) of the 4-quark contributions can be performed [44, 45] at low hadronic recoil for $q^2 \gtrsim (14 - 15) \text{ GeV}^2$ with the expansion parameter $\lambda = \Lambda/\sqrt{q^2} \sim \Lambda/m_b$. Moreover, form-factor relations [46–49] from the symmetries of QCD dynamics guide the construction of observables with reduced hadronic uncertainties in both kinematic regions.

A large amount of phenomenological studies considering form-factor symmetries have focused mainly on the decay $B \rightarrow K^*(\rightarrow K\pi)\bar{\ell}\ell$. The angular distribution of its 4-body final state [37, 50] comprises an order of ten observables that provide complementary information at low- and high- q^2 . Suitable combinations of these observables have also been identified that have either *i*) reduced hadronic uncertainties and possibly higher sensitivities to BSM contributions [51–61]; or *ii*) become short-distance independent, allowing one to gain information on form factors [57]. The decay $B \rightarrow K\bar{\ell}\ell$ offers fewer observables, some of which are sensitive to scalar and pseudoscalar [62, 63] interactions; and in the high- q^2 region the same short-distance dependence as in $B \rightarrow K^*\bar{\ell}\ell$ can be tested [64].

We perform a model-independent fit of the short-distance couplings $\mathcal{C}_{7,9,10}$ to the experimental data for exclusive $B \rightarrow K^*\gamma$, $B \rightarrow K^{(*)}\bar{\ell}\ell$, and $B_s \rightarrow \bar{\mu}\mu$ decays considering the standard set of operators described in more detail in Sec. 2. Improving on our previous analyses [57, 59, 64], we include the latest experimental data and add $B \rightarrow K^*\gamma$ and $B_s \rightarrow \bar{\mu}\mu$, as collected in Sec. 2. We go beyond [65] by including high- q^2 data for $B \rightarrow K^*\bar{\ell}\ell$ and the measurements of $B \rightarrow K\bar{\ell}\ell$; however, we do not consider inclusive measurements. In comparison to the very recent analysis [66], we use updated data and include $B \rightarrow K\bar{\ell}\ell$, but again do not consider the inclusive decays. For our analysis and for all numerical evaluations we use EOS [67].

Our analysis differs from all previous works [57, 59, 64–66] in its application of Bayesian inference with the help of Monte-Carlo techniques to treat theory uncertainties in the form of nuisance parameters. The statistical treatment and the choice of priors, as well as the determination of credibility intervals, goodness of fit, pull values, and Bayes factors for model comparison are described in Sec. 3. In Sec. 4, we present the results of the fit for short-distance couplings and discuss those nuisance parameters that are affected by data. We present updated predictions based on the fit results for unmeasured, optimized observables in the angular analysis of $B \rightarrow K^*(\rightarrow K\pi)\bar{\ell}\ell$ in the given scenario. The numerical input and details of the implementation of observables and nuisance parameters are summarized in App. A and App. B, respectively. We also present updated SM predictions in App. C. App. D contains definitions of distributions that have been used to model priors.

2 $\Delta B=1$ Decays: Conventions, Observables and Experimental Input

Rare $\Delta B=1$ decays are described by the effective theory of electroweak interactions. In the SM, the short-distance effects of heavy degrees of freedom of the order of the electroweak scale, due to the W and Z bosons and the t quark, are contained in the Wilson coefficients \mathcal{C}_i . The dynamics of the light-quark ($q = u, d, s, c, b$) and leptonic ($\ell = e, \mu, \tau$) degrees of freedom at the scale of the b quark are described by operators \mathcal{O}_i of dimension 5 and 6 for the parton transitions $b \rightarrow s + (\gamma, g, \bar{q}q, \bar{\ell}\ell)$. The SM Wilson coefficients \mathcal{C}_i ($i = 1, \dots, 10$) are presently known up to NNLO (and partially NNNLO) in QCD [28, 68–72] and NLO in QED [31, 32, 73, 74]. This includes the renormalization group evolution (RGE) from the electroweak scale $\mu_W \sim M_W$ down to $\mu_b \sim m_b$, the b -quark mass, which resums sizable logarithmic corrections to all orders in the QCD coupling α_s . Beyond the SM, the effects due to new heavy degrees of freedom can be included systematically as additional contributions to the short-distance couplings, possibly giving rise to operators beyond the SM with a different chiral nature or additional light degrees of freedom.

2.1 $\Delta B=1$ Effective Theory

The effective Hamiltonian of $\Delta B=1$ decays reads [28, 68]

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \left(\mathcal{H}_{\text{eff}}^{(t)} + \hat{\lambda}_u \mathcal{H}_{\text{eff}}^{(u)} \right), \quad \hat{\lambda}_u = V_{ub} V_{us}^* / V_{tb} V_{ts}^*, \quad (2.1)$$

$$\mathcal{H}_{\text{eff}}^{(t)} = \mathcal{C}_1 \mathcal{O}_1^c + \mathcal{C}_2 \mathcal{O}_2^c + \sum_{3 \leq i} \mathcal{C}_i \mathcal{O}_i, \quad \mathcal{H}_{\text{eff}}^{(u)} = \mathcal{C}_1 (\mathcal{O}_1^c - \mathcal{O}_1^u) + \mathcal{C}_2 (\mathcal{O}_2^c - \mathcal{O}_2^u) \quad (2.2)$$

where V_{ij} denotes an element of the Cabibbo-Kobayashi-Maskawa (CKM) quark-mixing matrix, and its unitarity relations have been used. Above and throughout, the Wilson coefficients are understood to be $\overline{\text{MS}}$ renormalized and taken at the reference scale $\mu = 4.2 \text{ GeV}$.¹ In the SM, all CP-violating effects in $b \rightarrow s$ transitions are governed by $\hat{\lambda}_u$ which is doubly Cabibbo suppressed and leads to tiny CP violation. The operators due to $b \rightarrow s \bar{q}q$ transitions are the current-current operators $\mathcal{O}_{1,2}^{u,c}$, the QCD-penguin operators for $i = 3, 4, 5, 6$, and the chromomagnetic dipole operator $i = 8$. Effects of QED-penguin operators are neglected since they are small for the decays under consideration. Following the studies of QED corrections to the inclusive decay, we choose the QED coupling α_e at the low scale μ_b , capturing most effects of QED corrections [31, 32] and removing the main uncertainty due to the choice of the renormalization scheme at LO in QED. The electromagnetic dipole operator

$$\mathcal{O}_7 = \frac{e}{(4\pi)^2} m_b [\bar{s} \sigma_{\mu\nu} P_R b] F^{\mu\nu} \quad (2.3)$$

governs $b \rightarrow s\gamma$ transitions. The semileptonic operators

$$\mathcal{O}_9 = \frac{\alpha_e}{4\pi} [\bar{s} \gamma_\mu P_L b] [\bar{\ell} \gamma^\mu \ell], \quad \mathcal{O}_{10} = \frac{\alpha_e}{4\pi} [\bar{s} \gamma_\mu P_L b] [\bar{\ell} \gamma^\mu \gamma_5 \ell] \quad (2.4)$$

¹Note that the actual low-energy renormalization scale μ_b might differ from μ , and the corresponding RGE effect $\mathcal{C}_i(\mu_b) = U(\mu_b, \mu)_{ij} \mathcal{C}_j(\mu)$ should be taken into account in renormalization-scale variations when determining the related uncertainty. Throughout we use a central value $\mu_b = \mu$.

observable	value	correlation	
$\mathcal{B} \times 10^5$	$4.55^{+0.72}_{-0.68} \pm 0.34$		[1]
	$4.47 \pm 0.10 \pm 0.16$		[7]
	$4.01 \pm 0.21 \pm 0.17$		[10]
S	$-0.03 \pm 0.29 \pm 0.03$	5%	[5]
C	$-0.14 \pm 0.16 \pm 0.03$		
S	$-0.32^{+0.36}_{-0.33} \pm 0.05$	8%	[12]
C	$+0.20 \pm 0.24 \pm 0.05$		

Table 1. Experimental results for CP-averaged $B^0 \rightarrow K^{*0} \gamma$ observables: branching fraction \mathcal{B} (CLEO, BaBar, Belle) and time-dependent CP asymmetries S and C (BaBar, Belle), including their correlations. Throughout, statistical errors are given first, followed by the systematic errors.

govern $b \rightarrow s \bar{\ell} \ell$ transitions, in combination with less important contributions from \mathcal{O}_7 .

In this study, we fit the Wilson coefficients $\mathcal{C}_{7,9,10}$ at the reference scale $\mu = 4.2$ GeV using experimental data. We assume them to be real valued and refrain from the frequently used decomposition into SM and BSM contributions $\mathcal{C}_i = \mathcal{C}_i^{\text{SM}} + \mathcal{C}_i^{\text{BSM}}$. The Wilson coefficients $i \leq 6$ and $i = 8$ contribute numerically only at the subleading level in the observables of interest and are fixed to the corresponding SM values at NNLO in QCD.

Whereas our scenario corresponds to the SM or extensions that do not introduce new CP violation nor new operators, more general scenarios have been investigated in the literature. The extension of this scenario with complex Wilson coefficients — i.e., CP violation beyond the SM — but no additional operators was studied in [59, 64, 66]. An extended operator basis with real Wilson coefficients, including chirality-flipped operators $i = 7', 9', 10'$, has been analyzed in [65, 66]. Finally, the combination of both can be found in [66]. Beyond these scenarios, it is conceivable that scalar, pseudoscalar, and tensor $b \rightarrow s \bar{\ell} \ell$ ($\ell = e, \mu$) operators can also contribute to the observables under consideration [54, 58, 63]. Beyond such direct contributions, additional ones can arise due to operator mixing from $b \rightarrow s \bar{q} q$ operators [75, 76] as well as $b \rightarrow s \bar{\tau} \tau$ [77].

2.2 Observables and Experimental Input

Phenomenological studies have analyzed and proposed a large number of CP-symmetric and -asymmetric observables. We summarize observables that either have been measured and therefore impose constraints on the Wilson coefficients or observables which are *i*) sensitive to the operators of interest and *ii*) exhibit a reduced hadronic uncertainty. For the latter, we compute the ranges that are still allowed by the data within the chosen scenario. Throughout, experimental numbers refer to CP-averaged quantities.

2.2.1 $B \rightarrow K^* \gamma$

For $B \rightarrow K^* \gamma$, several observables have been measured, such as the branching ratio \mathcal{B} , the time-dependent CP asymmetries S and C , and the isospin asymmetry A_I . Their impact on the scenario of real $\mathcal{C}_{7,7'}$ has been studied in [65] using the inclusive \mathcal{B} instead

q^2 -bin [GeV ²]	[1.00, 6.00]	[14.18, 16.00]	[> 16.00]	
$\langle \mathcal{B} \rangle \times 10^7$	$2.05_{-0.48}^{+0.53} \pm 0.07$	$1.46_{-0.36}^{+0.41} \pm 0.06$	$1.02_{-0.42}^{+0.47} \pm 0.06$	[8]
	$1.36_{-0.21}^{+0.23} \pm 0.08$	$0.38_{-0.12}^{+0.19} \pm 0.02$	$0.98_{-0.18}^{+0.20} \pm 0.06$	[13]
	$1.41 \pm 0.20 \pm 0.09$	$0.53 \pm 0.10 \pm 0.03$	$0.48 \pm 0.11 \pm 0.03$	[15]

Table 2. Experimental results for the CP-averaged branching fraction of charged $B^\pm \rightarrow K^\pm \bar{\mu} \mu$ decays from BaBar [8], Belle [13], and CDF [15], integrated in bins of q^2 . The publicly available results of BaBar and Belle are unknown admixtures of charged and neutral B decays. The difference between interpreting the data as coming from either purely charged or purely neutral B decays is negligible [64].

of the exclusive one. The measurement of $B_s \rightarrow \phi \gamma$ can provide similar information and allows a third CP asymmetry H to be studied [78]. The angular distribution in the decay $B \rightarrow K_1(1270)\gamma \rightarrow (K\pi\pi)\gamma$ is sensitive to the photon polarization and tests $\mathcal{C}_{7,7'}$; however, the feasibility of an analysis remains uncertain [79, 80]. In our analysis we use \mathcal{B} and the CP asymmetries S and C of $B \rightarrow K^* \gamma$ with their measurements and correlations compiled in Tab. 1, and follow the calculations outlined in [40, 81]. More details on the numerical input and nuisance parameters can be found in App. A and App. B.

2.2.2 $B \rightarrow K \bar{\ell} \ell$

In principle, the exclusive decay $B \rightarrow K \bar{\ell} \ell$ with a 3-body final state offers three (CP-averaged) observables: the branching ratio $\mathcal{B}(q^2)$, the lepton forward-backward asymmetry $A_{\text{FB}}(q^2)$, and the flat term $F_H(q^2)$. The latter two arise in the double-differential decay rate when differentiating with respect to the dilepton invariant mass q^2 and $\cos \theta_\ell$ [63]

$$\frac{1}{d\Gamma/dq^2} \frac{d^2\Gamma}{dq^2 d\cos \theta_\ell} = \frac{3}{4} (1 - F_H) \sin^2 \theta_\ell + \frac{1}{2} F_H + A_{\text{FB}} \cos \theta_\ell, \quad (2.5)$$

where θ_ℓ is the angle between the 3-momenta of the negatively charged lepton and the \bar{B} meson in the dilepton center of mass system. Two further interesting observables are the rate CP asymmetry A_{CP} and the ratio of decay rates for the $\ell=e$ and $\ell=\mu$ modes R_K . A_{FB} is nonzero only in the presence of scalar or tensor BSM contributions, and F_H is helicity suppressed by $m_\ell/\sqrt{q^2}$ in the scenario under consideration, but is sensitive to scalar and tensor contributions [62, 63]. In view of this, available measurements of A_{FB} , F_H , and R_K are not considered, and we include only the \mathcal{B} measurements for one low- q^2 and two high- q^2 bins as listed in Tab. 2. Our theory evaluation at low and high q^2 follows [63, 64]. Details concerning numerical input and nuisance parameters are given in App. A and App. B.

2.2.3 $B \rightarrow K^*(\rightarrow K\pi) \bar{\ell} \ell$

Phenomenologically, the angular analysis of the 4-body final state $B \rightarrow K^*(\rightarrow K\pi) \bar{\ell} \ell$ offers a large set of “angular” observables

$$\langle J_i \rangle [q_{\min}^2, q_{\max}^2] = \int_{q_{\min}^2}^{q_{\max}^2} dq^2 J_i(q^2), \quad i = 1, \dots, 9, \quad (2.6)$$

q^2 -bin [GeV ²]	[1.00, 6.00]	[14.18, 16.00]	[> 16.00]	
$\langle \mathcal{B} \rangle \times 10^7$	$2.05^{+0.53}_{-0.48} \pm 0.07$	$1.46^{+0.41}_{-0.36} \pm 0.06$	$1.02^{+0.47}_{-0.42} \pm 0.06$	[8]
	$1.49^{+0.45}_{-0.40} \pm 0.12$	$1.05^{+0.29}_{-0.26} \pm 0.08$	$2.04^{+0.27}_{-0.24} \pm 0.16$	[13]
	$1.42 \pm 0.41 \pm 0.08$	$1.34 \pm 0.26 \pm 0.08$	$0.97 \pm 0.26 \pm 0.06$	[15]
	$2.10 \pm 0.20 \pm 0.20$	$1.0738 \pm 0.1274 \pm 0.0728$	$1.32 \pm 0.15 \pm 0.09$	[18]
$\langle A_{\text{FB}} \rangle$	$-0.02^{+0.18}_{-0.16} \pm 0.07$	$-0.31^{+0.19}_{-0.11} \pm 0.13$	$-0.34^{+0.26}_{-0.17} \pm 0.08$	[9]
	$-0.26^{+0.30}_{-0.27} \pm 0.07$	$-0.70^{+0.22}_{-0.16} \pm 0.10$	$-0.66^{+0.16}_{-0.11} \pm 0.04$	[13]
	$-0.36^{+0.28}_{-0.46} \pm 0.11$	$-0.40^{+0.21}_{-0.18} \pm 0.07$	$-0.66^{+0.26}_{-0.18} \pm 0.19$	[16]
	$0.18 \pm 0.06^{+0.02}_{-0.01}$	$-0.49^{+0.06}_{-0.04} {}^{+0.05}_{-0.02}$	$-0.30 \pm 0.07 {}^{+0.01}_{-0.04}$	[18]
$\langle F_L \rangle$	$0.47 \pm 0.13 \pm 0.04$	$0.42^{+0.12}_{-0.16} \pm 0.11$	$0.47^{+0.18}_{-0.20} \pm 0.13$	[9]
	$0.67 \pm 0.23 \pm 0.05$	$-0.15^{+0.27}_{-0.23} \pm 0.07$	$0.12^{+0.15}_{-0.13} \pm 0.02$	[13]
	$0.60^{+0.21}_{-0.23} \pm 0.09$	$0.32 \pm 0.14 \pm 0.03$	$0.16^{+0.22}_{-0.18} \pm 0.06$	[16]
	$0.66 \pm 0.06^{+0.04}_{-0.03}$	$0.35^{+0.07}_{-0.06} {}^{+0.07}_{-0.02}$	$0.37^{+0.06}_{-0.07} {}^{+0.03}_{-0.04}$	[18]
$\langle A_T^{(2)} \rangle$	$1.6^{+1.8}_{-1.9} \pm 2.2$	$0.4 \pm 0.8 \pm 0.2$	$-0.9 \pm 0.8 \pm 0.4$	[16]
$\langle 2S_3 \rangle$	$0.10^{+0.15}_{-0.16} {}^{+0.02}_{-0.01}$	$0.04^{+0.15}_{-0.19} {}^{+0.04}_{-0.02}$	$-0.47^{+0.21}_{-0.10} {}^{+0.03}_{-0.05}$	[18]

Table 3. Experimental results used for $B^0 \rightarrow K^{*0} \bar{\ell} \ell$ for the CP-averaged branching fraction \mathcal{B} , lepton forward-backward asymmetry A_{FB} , longitudinal K^* -polarization fraction F_L , the transversity observable $A_T^{(2)}$ and $(2S_3)$ from BaBar [8, 9], Belle [13], CDF [15, 16], and LHCb [18]. Note that the sign of A_{FB} is reversed due to a different definition of θ_ℓ in the experimental community.

where the boundaries of the q^2 bin (throughout in units of GeV²) will not be explicitly shown when they are not relevant. Throughout, we assume that the experimental measurements are given for a certain q^2 binning that requires q^2 integration for theory predictions. Consequently, whenever a q^2 -dependent observable $X(q^2)$ is defined in a functional form $X(q^2) = f[J_i](q^2)$ in terms of the angular observables, we define the corresponding q^2 -integrated quantity as follows [57]

$$\langle X \rangle = f[\langle J_i \rangle]. \quad (2.7)$$

The angular observables $\langle J_i \rangle$ are defined in the 3-fold angular distribution

$$\begin{aligned} \frac{32\pi}{9} \frac{d^3\langle \Gamma \rangle}{d\cos\theta_\ell d\cos\theta_K d\phi} = & \\ & [\langle J_{1s} \rangle + \langle J_{2s} \rangle \cos 2\theta_\ell + \langle J_{6s} \rangle \cos \theta_\ell] \sin^2\theta_K + [\langle J_{1c} \rangle + \langle J_{2c} \rangle \cos 2\theta_\ell + \langle J_{6c} \rangle \cos \theta_\ell] \cos^2\theta_K \\ & + \langle J_3 \rangle \sin^2\theta_K \sin^2\theta_\ell \cos 2\phi + \langle J_4 \rangle \sin 2\theta_K \sin 2\theta_\ell \cos \phi + \langle J_5 \rangle \sin 2\theta_K \sin \theta_\ell \cos \phi \\ & + \langle J_7 \rangle \sin 2\theta_K \sin \theta_\ell \sin \phi + \langle J_8 \rangle \sin 2\theta_K \sin 2\theta_\ell \sin \phi + \langle J_9 \rangle \sin^2\theta_K \sin^2\theta_\ell \sin 2\phi, \end{aligned} \quad (2.8)$$

which accounts for all possible $(\bar{s} \dots b)(\bar{\ell} \dots \ell)$ Lorentz structures of chirality-flipped, scalar, pseudoscalar, and tensor operators [54, 58]. The angles are: *i*) θ_ℓ between the ℓ^- and the

K^* direction of flight in the $(\ell^+\ell^-)$ center of mass, *ii*) θ_K between the K and the K^* in the $(K\pi)$ center of mass and *iii*) ϕ between the $(\ell^+\ell^-)$ and $(K\pi)$ decay planes [51]. Here the normalization of the J_i from [51, 53, 54] is used and differs by a factor 4/3 from [52, 57, 59]. The following simplifications arise in the limit $m_\ell \rightarrow 0$ and in the absence of scalar and tensor operators [54, 58]:

$$J_{1s} = 3 J_{2s}, \quad J_{1c} = -J_{2c}, \quad J_{6c} = 0, \quad (2.9)$$

and a fourth more complicated relation [56]. It is straightforward to obtain the decay rate and the three single-differential angular distributions from (2.8)

$$\langle \Gamma \rangle = \frac{3}{4} [2\langle J_{1s} \rangle + \langle J_{1c} \rangle] - \frac{1}{4} [2\langle J_{2s} \rangle + \langle J_{2c} \rangle], \quad (2.10)$$

$$\frac{d\langle \Gamma \rangle}{d\phi} = \frac{1}{2\pi} [\langle \Gamma \rangle + \langle J_3 \rangle \cos 2\phi + \langle J_9 \rangle \sin 2\phi], \quad (2.11)$$

$$\frac{d\langle \Gamma \rangle}{d\cos \theta_K} = \frac{3}{8} [(3\langle J_{1s} \rangle - \langle J_{2s} \rangle) \sin^2 \theta_K + (3\langle J_{1c} \rangle - \langle J_{2c} \rangle) \cos^2 \theta_K], \quad (2.12)$$

$$\frac{d\langle \Gamma \rangle}{d\cos \theta_\ell} = \frac{3}{8} [2\langle J_{1s} \rangle + \langle J_{1c} \rangle + (2\langle J_{6s} \rangle + \langle J_{6c} \rangle) \cos \theta_\ell + (2\langle J_{2s} \rangle + \langle J_{2c} \rangle) \cos 2\theta_\ell]. \quad (2.13)$$

The branching ratio $\langle \mathcal{B} \rangle$, the lepton forward-backward asymmetry $\langle A_{\text{FB}} \rangle$, and the longitudinal K^* -polarization fraction $\langle F_L \rangle$

$$\langle \mathcal{B} \rangle = \tau_{B^0} \langle \Gamma \rangle, \quad \langle A_{\text{FB}} \rangle = \frac{3}{8} \frac{2\langle J_{6s} \rangle + \langle J_{6c} \rangle}{\langle \Gamma \rangle}, \quad \langle F_L \rangle = \frac{3\langle J_{1c} \rangle - \langle J_{2c} \rangle}{4\langle \Gamma \rangle}, \quad (2.14)$$

have been measured by BaBar [8, 9], Belle [13], CDF [15, 16], and LHCb [18]. The angular observable $\langle A_T^{(2)} \rangle$ [51] has been measured by CDF [16]; and $\langle S_3 \rangle$ [54] has been determined by LHCb [18]:

$$\langle A_T^{(2)} \rangle = \frac{\langle J_3 \rangle}{2\langle J_{2s} \rangle}, \quad \langle S_3 \rangle = \frac{\langle J_3 \rangle}{\langle \Gamma \rangle}. \quad (2.15)$$

All are summarized in Tab. 3. Note that $\langle A_{\text{FB}} \rangle$ and $\langle F_L \rangle$ are determined from a combined fit to the single-differential angular distributions

$$\frac{1}{\langle \Gamma \rangle} \frac{d\langle \Gamma \rangle}{d\cos \theta_K} = \frac{3}{4} [1 - \langle F_L \rangle] \sin^2 \theta_K + \frac{3}{2} \langle F_L \rangle \cos^2 \theta_K, \quad (2.16)$$

$$\frac{1}{\langle \Gamma \rangle} \frac{d\langle \Gamma \rangle}{d\cos \theta_\ell} = \frac{3}{4} \langle F_L \rangle \sin^2 \theta_\ell + \frac{3}{8} [1 - \langle F_L \rangle] (1 + \cos^2 \theta_\ell) + \langle A_{\text{FB}} \rangle \cos \theta_\ell. \quad (2.17)$$

The observables $\langle A_T^{(2)} \rangle$ and $\langle A_{\text{im}} \rangle = \langle J_9 \rangle / \langle \Gamma \rangle$ are determined from

$$\frac{2\pi}{\langle \Gamma \rangle} \frac{d\langle \Gamma \rangle}{d\phi} = 1 + \frac{1}{2} [1 - \langle F_L \rangle] \langle A_T^{(2)} \rangle \cos 2\phi + \langle A_{\text{im}} \rangle \sin 2\phi, \quad (2.18)$$

implying $S_3 = (1 - \langle F_L \rangle) \langle A_T^{(2)} \rangle / 2$. Note that (2.17) and (2.18) are based on the approximation (2.9), which is well justified within our scenario.

The angular observables $\langle J_i \rangle$ and the branching ratio $\langle \mathcal{B} \rangle$ are proportional to the square of hadronic form factors, the main source of theory uncertainty. In normalized combinations of the angular observables, for example A_{FB} and F_L , these uncertainties partially cancel. The most prominent example is the position $q_0^2[A_{\text{FB}}]$ of the zero crossing of A_{FB} . A number of suitable combinations have been found for both low- and high- q^2 regions. At low q^2 [51, 53, 56, 60, 61]

$$\langle A_T^{(2)} \rangle = \frac{\langle J_3 \rangle}{2 \langle J_{2s} \rangle}, \quad \langle A_T^{(\text{re})} \rangle = \frac{\langle J_{6s} \rangle}{4 \langle J_{2s} \rangle}, \quad \langle A_T^{(\text{im})} \rangle = \frac{\langle J_9 \rangle}{2 \langle J_{2s} \rangle}, \quad (2.19)$$

$$\langle A_T^{(3)} \rangle = \sqrt{\frac{\langle 2 J_4 \rangle^2 + \langle J_7 \rangle^2}{-2 \langle J_{2c} \rangle \langle 2 J_{2s} + J_3 \rangle}}, \quad \langle A_T^{(4)} \rangle = \sqrt{\frac{\langle J_5 \rangle^2 + \langle 2 J_8 \rangle^2}{\langle 2 J_4 \rangle^2 + \langle J_7 \rangle^2}}, \quad (2.20)$$

$$\langle A_T^{(5)} \rangle = \frac{\sqrt{\langle 4 J_{2s} \rangle^2 - \langle J_{6s} \rangle^2 - 4(\langle J_3 \rangle^2 + \langle J_9 \rangle^2)}}{8 \langle J_{2s} \rangle}; \quad (2.21)$$

whereas at high q^2 [57]

$$\langle H_T^{(1)} \rangle = \frac{\sqrt{2} \langle J_4 \rangle}{\sqrt{-2 \langle J_{2c} \rangle \langle 2 J_{2s} - J_3 \rangle}}, \quad (2.22)$$

$$\langle H_T^{(2)} \rangle = \frac{\langle J_5 \rangle}{\sqrt{-2 \langle J_{2c} \rangle \langle 2 J_{2s} + J_3 \rangle}}, \quad \langle H_T^{(3)} \rangle = \frac{\langle J_{6s} \rangle}{2 \sqrt{\langle 2 J_{2s} \rangle^2 - \langle J_3 \rangle^2}}. \quad (2.23)$$

For brevity, factors of $\beta_\ell = \sqrt{1 - 4 m_\ell^2 / q^2}$ have been set to unity, since they are negligible in our scenario for the considered range $q^2 \gtrsim 1 \text{ GeV}^2$. Recently, it was found that $H_T^{(1)}$ and $H_T^{(2)}$ are also optimized observables at low q^2 [61].

We note that at low q^2 , J_3 and J_9 vanish at leading order in QCDF [52], making them ideal probes of chirality-flipped operators $i = 7', 9', 10'$ because leading terms in QCDF are $\sim \text{Re}[\mathcal{C}_i \mathcal{C}_{i'}^*]$ and $\sim \text{Im}[\mathcal{C}_i \mathcal{C}_{i'}^*]$. J_9 (and also $J_{7,8}$) vanishes for real Wilson coefficients, and therefore the measurements of $\langle A_T^{(\text{im})} \rangle$ and $\langle A_{\text{im}} \rangle$ are not of interest for our scenario. Only partial results of the subleading corrections exist [40, 81] and only those of kinematic origin are included in the numerical evaluation. Nevertheless, $\langle A_T^{(2)} \rangle$ and $\langle 2S_3 \rangle$ are included in our fit because they might allow us to obtain information on the nuisance parameters used to model yet-unknown subleading contributions (see App. B.3).

At high q^2 , F_L and $A_T^{(2)}$ become short-distance independent [57] and the experimental data allow us to constrain the form-factor-related nuisance parameters; see App. B.2. This has been exploited recently [82] to extract the q^2 dependence of form factors from data and, comparing with preliminary lattice results, to find overall agreement within the still large uncertainties.

In our predictions, we therefore focus on the yet-unmeasured optimized observables $\langle A_T^{(\text{re},3,4,5)} \rangle$ and $\langle H_T^{(1,2,3)} \rangle$.

2.2.4 $B_s \rightarrow \bar{\mu}\mu$

The rare decay $B_s \rightarrow \bar{\mu}\mu$ is helicity suppressed in the SM, making it an ideal probe of contributions from scalar and pseudoscalar operators. Its branching ratio depends only on \mathcal{C}_{10} in the scenario under consideration

$$\mathcal{B}[B_s(t=0) \rightarrow \bar{\mu}\mu] = \frac{G_F^2 \alpha_e^2 M_{B_s}^3 f_{B_s}^2 \tau_{B_s}}{64 \pi^3} |V_{tb} V_{ts}^*|^2 \sqrt{1 - \frac{4m_\mu^2}{M_{B_s}^2}} \frac{4m_\mu^2}{M_{B_s}^2} |\mathcal{C}_{10}|^2 \quad (2.24)$$

and is predicted in the SM to be around 3×10^{-9} . The main uncertainties are due to the decay constant f_{B_s} and the CKM factor $|V_{tb} V_{ts}^*|$.

Above the mixing of B_s -meson has not been taken into account, i.e., the branching ratio refers to time $t = 0$. However, experimentally the time-integrated branching ratio is determined. Both are related in our SM-like scenario as [83]

$$\mathcal{B}[B_s \rightarrow \bar{\mu}\mu] = \frac{1}{1 - y_s} \mathcal{B}[B_s(t=0) \rightarrow \bar{\mu}\mu], \quad y_s = \frac{\Delta\Gamma_s}{2\Gamma_s}. \quad (2.25)$$

Lately, the most precise measurement of the life-time difference $\Delta\Gamma_s$ became available from LHCb [84] and moreover LHCb succeeded to determine the sign of $\Delta\Gamma_s$ [85] which turned out to be SM-like. In view of this, we will use the numerical value from LHCb $y_s = 0.088 \pm 0.014$ [84].

In the last decade, the Tevatron experiments DØ [19] and CDF [20, 21] lowered the upper bound on the branching ratio by several orders of magnitude to a value close to 1×10^{-8} ; and CDF announced the first direct evidence based on a 2σ fluctuation over the background-only hypothesis [20, 21]. This year the LHC experiments LHCb, CMS, and ATLAS provided their results based on the complete 2011 run [22–26]. In our analysis we use the most stringent result $\mathcal{B}(B_s \rightarrow \bar{\mu}\mu) < 4.5 \times 10^{-9}$ (3.8×10^{-9}) at 95% (90%) CL, obtained by LHCb [23]. Details of the implementation of this bound are given in Sec. 3.3.

3 Statistical Method

We have decided to use the full Bayesian approach in this analysis. It allows us to incorporate all available experimental results, to obtain probability statements about the parameters of interest $\vec{\theta}$ — the Wilson coefficients — and to compare different models using the Bayes factor.

In the Bayesian approach, we describe theory uncertainties by adding nuisance parameters $\vec{\nu}$. It is straightforward to incorporate existing knowledge — say from power counting, symmetry arguments, or even other dedicated Bayesian analyses — about these theory uncertainties by specifying informative priors. As a cross validation, it is useful to employ different priors and compare the posterior inference. Any significant discrepancy based on two different prior choices implies that more accurate experimental or theoretical input is needed before conclusive statements can be made. This can be seen as a feature of the Bayesian methodology. Throughout, we assume that parameters are independent a priori,

$$P(\vec{\theta}, \vec{\nu}) = \prod_i P(\theta_i) \cdot \prod_j P(\nu_j). \quad (3.1)$$

The experimental data D are used in the likelihood $P(D|\vec{\theta}, \vec{\nu})$, and Bayes' theorem yields the posterior knowledge about the parameters after learning from the data D

$$P(\vec{\theta}, \vec{\nu}|D) = \frac{P(D|\vec{\theta}, \vec{\nu}) P(\vec{\theta}, \vec{\nu})}{Z}, \quad (3.2)$$

with the normalization given by the evidence

$$Z = \int d\vec{\theta} d\vec{\nu} P(D|\vec{\theta}, \vec{\nu}) P(\vec{\theta}, \vec{\nu}). \quad (3.3)$$

In case we want to remove the dependence on $\vec{\nu}$ in the posterior, we simply marginalize:

$$P(\vec{\theta}|D) = \int d\vec{\nu} P(\vec{\theta}, \vec{\nu}|D). \quad (3.4)$$

The integrations are performed with the Monte Carlo algorithm described next.

3.1 Monte Carlo Algorithm

The presence of multiple, well separated modes, the large dimensionality of the parameter space, and the costly evaluation of the likelihood require a sophisticated algorithm [86]. We sketch the main steps of this new algorithm:

1. A sufficiently large number of Markov chains are run in parallel for $\mathcal{O}(50000)$ iterations to explore the parameter space with an adaptive local random walk. The chains need a burn-in phase, thus we discard the first 15% of the iterations.
2. Chains whose common R -value [87] is reasonably small, say $R < 2$, are combined into groups.
3. We create patches with a length of $\mathcal{O}(1000)$ points from the individual chains and define a multivariate density from the mean and covariance of each patch.
4. Using hierarchical clustering [88], we combine the patches into a smaller number of clusters. As the initial guess for the clustering, we construct a fixed number of about 30 patches of length $\mathcal{O}(5000)$ from each group of chains.
5. We define a multivariate mixture density from the output of the clustering by assigning equal weights to each cluster. This mixture density serves as the initial proposal density for the Population Monte Carlo (PMC) algorithm [89, 90].
6. Using a computing cluster with a few hundred cores, we draw importance samples and adapt the proposal density to the posterior until convergence is achieved; i.e., until the difference in perplexity between two consecutive steps is less than 2%.
7. Given the resulting proposal density, we collect $2 \cdot 10^6$ importance samples to compute marginal distributions and the evidence.

The biggest advantages of this approach are the automatic adaptation to the complicated posterior shape, and the ability to massively parallelize the costly evaluation of the likelihood.

3.2 Priors

We use flat priors for the Wilson coefficients. This is not done because we want to imply complete prior ignorance, but, instead, we want a convenient, sufficiently diffuse density, with the expectation that the posterior is dominated by the likelihood.

For the nuisance parameters, the choice of prior depends on the parameter’s nature. There are the four quark-mixing matrix (CKM) parameters, the b and c quark masses, the decay constant f_{B_s} entering $B_s \rightarrow \bar{\mu}\mu$, and most dominantly the $B \rightarrow K^{(*)}$ form factors. In addition, unknown subleading contributions in the two distinct Λ/m_b expansions at large and low recoil are parametrized as nuisance parameters. The complete list of almost 30 nuisance parameters along with the choice of the prior densities is presented in App. B. Note that most nuisance parameters only affect a subset of the observables.

Where possible, the posterior distributions of the nuisance parameters from previous analyses fitting different data are used as the prior distributions in our fit. As an example, we use the output for the quark masses [91] in the form of LogGamma distributions; see App. D.2.

For the CKM parameters λ , A , $\bar{\rho}$, $\bar{\eta}$, we choose the results of the UTfit Collaboration [92]. When allowing for BSM contributions in our fit, we use the results of the so-called CKM *tree-level fit* for λ , A , $\bar{\rho}$, $\bar{\eta}$. The tree-level fit represents only the basic constraints from SM tree-level processes, which every extension of the SM must include and we assume that BSM contributions are negligible. Thus no information from rare B decays and B – \bar{B} mixing is used indirectly through the priors. For the SM predictions, we use the results of the SM-CKM fit instead. The posterior distributions from either fit are assumed to be symmetric Gaussian distributions, which was found to be a good approximation.

We model theory uncertainties with Gaussian distributions in cases where authors only report an estimate of the magnitude. This is justified by the principle of maximum entropy [93]. As an example, suppose the quoted uncertainty of a QCD form factor f is 15%. We introduce a nuisance parameter, ζ_f , as a scaling factor, such that $f \rightarrow \zeta_f \cdot f$, and we vary $\zeta_f \sim \mathcal{N}(\mu = 1, \sigma = 0.15)$, with an allowed range $\zeta_f \in [1 - 3 \cdot \sigma, 1 + 3 \cdot \sigma]$; i.e., neglecting the tails of the Gaussian beyond 3σ . If necessary, we modify the range to avoid unphysical values of f . Subleading phases are incorporated with flat priors covering the full range.

3.3 Experimental Results

We form the total log likelihood, $\log P(D|\vec{\theta}, \vec{\nu})$, by summing over the individual contributions. The complete list of experimental results used is given in tables 1 to 3. The majority of results is incorporated as 1-dimensional Gaussian distributions, whose variances are obtained by adding statistical and systematic uncertainties in quadrature, $\sigma^2 = \sigma_{stat}^2 + \sigma_{syst}^2$. In the case of asymmetric uncertainties, we use a piecewise function constructed from two Gaussian distributions around the central value with different variances. With the exception of the upper bound on $\mathcal{B}(B_s \rightarrow \bar{\mu}\mu)$, we limit ourselves to the Gaussian distribution in the likelihood, despite the discontinuities arising from asymmetric uncertainties, to obtain results that are comparable with the existing literature. Known correlations between observations, e.g., between the time-dependent CP asymmetries S and C in $B \rightarrow K^*\gamma$,

are represented by multivariate Gaussian distributions. Note that in $B \rightarrow K^* \bar{\ell} \ell$ decays the observables A_{FB} and F_L are extracted from a simultaneous fit to the double differential decay rate. Requiring physical values, i.e.,

$$\frac{d^2\Gamma}{dq^2 d\cos\theta_{\ell,K}} > 0, \quad (3.5)$$

cuts out an allowed region in the (A_{FB}, F_L) - plane. Since the fit for (A_{FB}, F_L) typically converges near the unphysical region, the resulting contribution to the likelihood would be distinctly non-Gaussian. Unfortunately, the resulting 2D likelihood is not publicly available, thus we assume A_{FB} and F_L independent and Gaussian distributed.

We include the results of direct searches for the decay $B_s \rightarrow \bar{\mu} \mu$ into the likelihood. Often, only the 90% and 95% limits on the branching ratio \mathcal{B} , obtained with the CL_S method [94], are published. However, there is no single best way to translate these limits into a useful contribution to the likelihood, and several schemes of varying sophistication exist in the literature [95–97]. It is preferable to directly use the Bayesian posterior on the branching ratio $P(\mathcal{B}|D)$, computed by a general algorithm for multichannel search experiments [98]. This posterior is often produced to compute Bayesian limits for cross checks with CL_S results. The input numbers — expected signal yields, background yields — that are needed to compute $P(\mathcal{B}|D)$ are publicly available from LHCb [23]; only the correlations of the yields are not published. By reinterpreting the function $P(\mathcal{B}|D)$ as $P(\mathcal{B}|\vec{\theta}, \vec{\nu})$ the desired contribution to the likelihood is found. For a convenient approximation to $P(\mathcal{B}|D)$, we use the four-parameter Amoroso-distribution [99] with relevant details given in App. D.1. For the data supplied by LHCb [23], the relative error of this interpolation is at most 2%.

3.4 Uncertainties of Theory Predictions

Within the Bayesian framework, the procedure to calculate the uncertainty of an observable’s prediction within a given theory, say the SM, is essentially uncertainty propagation. In this case, an observable A depends on Wilson coefficients and on additional nuisance parameters. We fix the values of the Wilson coefficients, $\vec{\theta} = \vec{\theta}_{\text{SM}}$, so the value of A is uniquely determined by $\vec{\nu}$; i.e., $A = f(\vec{\nu})$. We vary the nuisance parameters according to their prior, $P(\vec{\nu})$. The distribution of the random variable A , $P(A)$, is given by

$$P(A) = \int d\vec{\nu} P(A, \vec{\nu}) = \int d\vec{\nu} P(A|\vec{\nu}) P(\vec{\nu}) = \int d\vec{\nu} \delta(A - f(\vec{\nu})) P(\vec{\nu}), \quad (3.6)$$

where we used the Dirac δ -distribution. Numerically, one only needs to draw parameter samples $\vec{\nu}_i \sim P(\vec{\nu})$ and calculate A for each sample $\vec{\nu}_i$. We collect the resulting samples A_i in one dimensional histograms to extract 68% intervals. As before, we assume $P(\vec{\nu}) = \prod_j P(\nu_j)$ and use the priors listed in App. B. As a welcome side effect, this form of $P(\vec{\nu})$ allows us to efficiently sample $\vec{\nu}$ from the joint prior by sampling from simple, 1D priors directly without the need to resort to MCMC or PMC.

If we take this approach one step further, we can ask: what are the likely values, or informally speaking “the allowed ranges,” of A , given the set of measurements D listed in

Section 2? Using the full posterior on both Wilson coefficients and nuisance parameters, $P(\vec{\theta}, \vec{\nu} | D)$, we obtain

$$P(A | D) = \int d\vec{\theta} d\vec{\nu} \delta(A - f(\vec{\theta}, \vec{\nu})) P(\vec{\theta}, \vec{\nu} | D). \quad (3.7)$$

We simply take the posterior samples produced by the PMC algorithm, compute A_i for each sample $(\vec{\theta}, \vec{\nu})_i$, and finally fill the sample A_i with its importance weight into a histogram.

3.5 Goodness of Fit and Model Comparison

To check that the assumed model with three real Wilson coefficients provides a good description of the experimental observations, we determine the goodness of fit. We follow the standard procedure: first we choose a test statistic $T(D | \vec{\theta}, \vec{\nu})$ with the parameter values chosen at a local mode of the posterior, then calculate its distribution, and finally determine the value of the test statistic for the actual data set. For more details on p -values and how we interpret them in this work, we refer to [100]. We make two closely related choices for T , defined as follows.

For each observable x , we compare its theory prediction $x_{pred}(\vec{\theta}, \vec{\nu})$ with the mode of the experimental distribution (central value) of x , denoted by x^* . We compute the frequency f that a value of x less extreme than x_{pred} would be observed. Using the inverse of the Gaussian cumulative distribution function, $\Phi^{-1}(\cdot)$, we define the pull:

$$\delta = \Phi^{-1} \left[\frac{f + 1}{2} \right]. \quad (3.8)$$

Note that for a 1-dimensional Gaussian, this reduces to the usual $\delta = (x^* - x_{pred})/\sigma$. In the 1-dimensional case, the (Gaussian, Amoroso) distributions yield a signed δ (positive if $x^* > x_{pred}$, else negative), while for the multivariate Gaussian, δ is positive semidefinite. We define the test statistic T_{pull} as

$$T_{\text{pull}} = \sum_i \delta_i^2, \quad (3.9)$$

where i extends over all experimental data. As a cross check, we also consider T_{like} , defined as the value of the log likelihood, $T_{\text{like}} = \log P(D | \vec{\theta}, \vec{\nu})$. Its frequency distribution is approximated by generating 10^5 pseudo experiments $D \sim P(D | \vec{\theta}^*, \vec{\nu}^*)$, where $\vec{\theta}^*, \vec{\nu}^*$ are fixed values at a local maximum of the posterior. Since we do not have the raw data — events, detector simulations etc. — available, we generate pseudo experiments. Consider the case of a single measurement with Gaussian uncertainties $\mathcal{N}(\mu = x^*, \sigma)$: we fix the theory prediction, shifting the maximum of the Gaussian to $x_{pred}(\vec{\theta}^*, \vec{\nu}^*)$, but keep the uncertainties reported by the experiment. Then we generate $x \sim \mathcal{N}(\mu = x_{pred}(\vec{\theta}^*, \vec{\nu}^*), \sigma)$, and proceed analogously for all observables to sample D . The p -value is computed by counting the fraction of experiments with a likelihood value smaller than that for the observed data set and corrected for the number of degrees of freedom; see Section III.D.5 in [100]. Although the generation of pseudo data is far from perfect, we emphasize that, on the one hand, it is fast and, on the other hand, we will not consider the actual value of

p too rigorously. Two models with p -values of 40% and 60% both describe the data well, and that is all the information we need.

If we used the maximum likelihood parameters and ignored the $B_s \rightarrow \bar{\mu}\mu$ contribution, both statistics would be equivalent to χ^2 and thus yield the same p -value. The parameter values at the global mode of the posterior differ only little from the maximum likelihood values, and $B_s \rightarrow \bar{\mu}\mu$ presents only one of 59 inputs. We therefore consider it reasonable to approximate the distribution of T_{pull} by the χ^2 -distribution in order to compute a p -value.

If there are several local modes with reasonably high p -values, it is necessary to assess which of them is favored by the data; i.e., to perform a model comparison. Suppose the full parameter space is decomposed into disjoint subsets V_i , $i = 1 \dots n$, where V_i contains only a single mode of the posterior. Then we compute the local evidence Z_i by integrating over V_i in (3.3). In fact, the integral is available as the average weight of all importance samples in V_i , with an accuracy of roughly 5%. The Bayes factor $B_{ij} = Z_i/Z_j$ is the data-dependent part in the posterior odds of two statistical models M_i, M_j :

$$\frac{P(M_i|D)}{P(M_j|D)} = B_{ij} \cdot \frac{P(M_i)}{P(M_j)} \quad (3.10)$$

In addition to the local evidence, we compute the evidence Z_{SM} for the SM case with fixed values for the Wilson coefficients, but with $\bar{\nu}$ allowed to vary. Computing the Bayes factor with Z_{SM} and the local evidence for the region with SM-like signature allows us to assess if the data are in favor of adding three degrees of freedom for the Wilson coefficients to achieve a better agreement with the theory predictions, or if the SM is preferred due to its simplicity.

4 Results

In the following, we discuss the results of our analysis of the experimental data described in Sec. 2.2 using the statistical tools explained in Sec. 3. First, the results of the global fit of the three Wilson coefficients and 28 nuisance parameters to 59 experimental inputs are presented, including marginal distributions and best fit points. As for goodness of fit, we list p -values and evidence for each of the arising solutions. Furthermore, we show pull values for the included measurements. And we discuss the fit results of nuisance parameters, if the posterior differs significantly from the prior. Second, we provide predictions of yet-unmeasured, optimized observables in $B \rightarrow K^*(\rightarrow K\pi)\bar{\ell}\ell$ at low and high q^2 within our scenario, taking into account the experimental data. Finally, we give SM predictions for measured and yet-unmeasured observables including theory uncertainties determined using the Monte Carlo method as explained in Sec. 3.4.

4.1 Fit Results

Here, we summarize the main part of our work: the results of the fits of the parameters of interest, the Wilson coefficients $\mathcal{C}_{7,9,10}$, to the data listed in Sec. 2.2. Details of the Monte Carlo algorithm are given in Sec. 3.1. The treatment of prior distributions is explained in

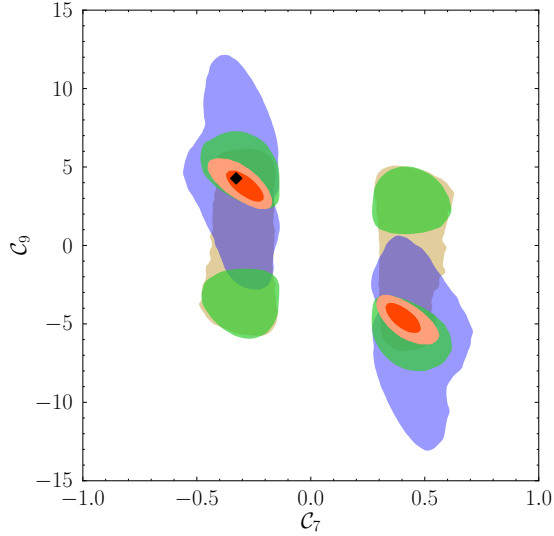
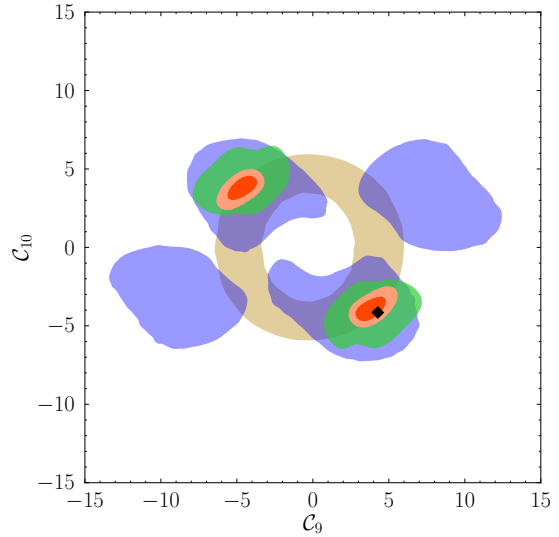
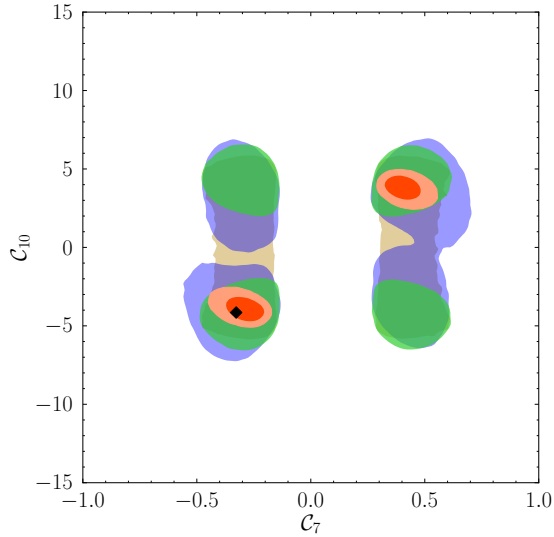


Figure 1. The marginalized 2-dimensional 95% credibility regions of the Wilson coefficients $\mathcal{C}_{7,9,10}$ for $\mu = 4.2$ GeV are shown when applying the $B \rightarrow K^* \gamma$ constraints in combination with *i)* only low- and high- q^2 data from $B \rightarrow K \bar{\ell} \ell$ [brown]; *ii)* only low- q^2 data from $B \rightarrow K^* \bar{\ell} \ell$ [blue]; *iii)* only high- q^2 data from $B \rightarrow K^* \bar{\ell} \ell$ [green]; and *iv)* all the data, including also $B_s \rightarrow \bar{\mu} \mu$ [light red], showing as well the 68% credibility interval [red]. The SM values $\mathcal{C}_{7,9,10}^{\text{SM}}$ are indicated by \blacklozenge .



Sec. 3.2, the priors of the nuisance parameters are specified in App. B. For $\mathcal{C}_{7,9,10}$, we use flat priors with

$$\mathcal{C}_7 \in [-1, 1], \quad \mathcal{C}_{9,10} \in [-10, 10]. \quad (4.1)$$

The fit not only constrains the Wilson coefficients $\mathcal{C}_{7,9,10}$, but updates our knowledge of the nuisance parameters. We discuss the most significant changes.

The marginalized two-dimensional 95% credibility regions are shown in Fig. 1 when applying the $B \rightarrow K^* \gamma$ constraints (Sec. 2.2.1) in combination with *i)* only low- and high- q^2 data from $B \rightarrow K \bar{\ell} \ell$ (Sec. 2.2.2); *ii)* only low- q^2 data from $B \rightarrow K^* \bar{\ell} \ell$ ²; *iii)* only high- q^2 data from $B \rightarrow K^* \bar{\ell} \ell$ (Sec. 2.2.3); and finally *iv)* all the data, including also $B_s \rightarrow \bar{\mu} \mu$ (Sec. 2.2.4).

²Here we enlarged the prior ranges of $\mathcal{C}_{7,9,10}$ by a factor of 2, which is irrelevant for the remainder of our work.

The most stringent constraints on $\mathcal{C}_{9,10}$ come from the high- q^2 data of $B \rightarrow K^* \bar{\ell} \ell$, which should be taken with some caution since the form factors are only available as extrapolations of LCSR results from low q^2 . In the near future, we expect more accurate lattice calculations of form factors to close this weak point. Also shown are the SM predictions of $\mathcal{C}_{7,9,10}(\mu = 4.2 \text{ GeV})$ using NNLO evolution [31].

We confirm the findings of previous analyses [64, 66] that only two solutions make up 95% of the probability: the first exhibits the same signs of $\mathcal{C}_{7,9,10}$ as the SM. The second solution corresponds to a first order degeneracy of all observables under a simultaneous sign flip $\mathcal{C}_{7,9,10} \rightarrow -\mathcal{C}_{7,9,10}$. There are two additional local maxima that correspond to a sign flip of $\mathcal{C}_7 \rightarrow -\mathcal{C}_7$ of the former solutions. In Tab. 4, we list the properties of these four modes, categorized by the signs of $\mathcal{C}_{7,9,10}$. As witnessed by the evidence, Z , the SM-like and sign-flipped solutions essentially make up the whole posterior mass, with ratios of 52% and 48%, respectively. The other two solutions are suppressed by many orders of magnitude, and thus do not appear at the 95% level. For the two dominant solutions, the goodness-of-fit results are nearly identical: both p -values based on the statistics T_{like} and T_{pull} (see Sec. 3.5) are large, indicating a good fit. In contrast, the suppressed solutions do not seem to explain the data well. We note that the MCMC revealed a handful of additional modes with $6 \lesssim |\mathcal{C}_{9,10}| \lesssim 9$. We do not consider these further because they are suppressed by a factor of roughly $\exp(40)$ compared to the global maximum.

To highlight the fit results, we present the pull (3.8) — the normalized deviation between theory prediction and measured value in units of Gaussian σ — for all 59 constraints. Pulls for $B \rightarrow K^* \gamma$ [left] and $B \rightarrow K \bar{\ell} \ell$ [right] constraints are shown in Fig. 2; those for $B \rightarrow K^* \bar{\ell} \ell$ in Fig. 3; and the pull for LHCb’s result of $B_s \rightarrow \bar{\mu} \mu$ is -1.1 ; i.e., its most likely value from the measurement is about 1σ (in terms of the experimental uncertainty) lower than the theory prediction. Here, the theory parameters are chosen at the global maximum of the posterior. With the best-fit parameters with $\mathcal{C}_{7,9,10}$ fixed at SM values (see below), we obtain nearly identical plots, and we therefore omit them. We observe the largest pull at $+2.5$ for the Belle measurement of $\langle \mathcal{B} \rangle$ [16, 19, 21] for $B \rightarrow K^* \bar{\ell} \ell$. It is the only pull surpassing 2.0. Fig. 3 shows, for example, how the debate about the existence of a zero crossing of A_{FB} at large recoil was settled: the first published measurements by Belle and CDF deviated from the SM prediction, but when taken together with LHCb’s recent result that pulls the best fit point towards the SM, there is good agreement between the SM and the experiments.

We also perform the global fit with $\mathcal{C}_{7,9,10}$ fixed to the SM values, varying only the nuisance parameters; see the bottom row in Tab. 4. The prior normalization then changes by $\log(800) = 6.68$ due to omitting $\mathcal{C}_{7,9,10}$ with ranges given in (4.1). The values of T_{like} and T_{pull} are just as good as for the two dominant solutions, but the p -values are even larger, as the number of degrees of freedom used in the χ^2 -distribution to calculate p differs by three. We compute the Bayes factor of the SM vs the SM-like solution by dividing their respective evidences:

$$B = \exp(392.6 - 385.3) \approx 1500. \quad (4.2)$$

Assuming prior odds of one, the posterior odds are given by B , and thus are clearly

$\text{sgn}(\mathcal{C}_7, \mathcal{C}_9, \mathcal{C}_{10})$	best-fit point	$\log(\text{MAP})$	goodness of fit				$\log(Z)$
			T_{like}	p_{like}	T_{pull}	p_{pull}	
$(-, +, -)$	$(-0.293, 3.69, -4.19)$	425.22	402.59	60%	48.4	75%	385.3
$(+, -, +)$	$(0.416, -4.59, 4.05)$	425.08	402.49	60%	48.5	75%	385.2
$(-, -, +)$	$(-0.393, -3.12, 3.20)$	404.67	387.88	0.9%	76.5	4%	363.9
$(+, +, -)$	$(0.558, 2.25, -3.24)$	400.91	384.52	0.2%	83.1	1%	358.9
SM: $(-, +, -)$	$(-0.327, 4.28, -4.15)$	431.46 [†]	402.53	70%	48.5	83%	392.6

Table 4. Best-fit point, log maximum a-posteriori (MAP) value, goodness of fit summary and log evidence for the four local modes (denoted by the signs of $(\mathcal{C}_7, \mathcal{C}_9, \mathcal{C}_{10})$) of the posterior including all experimental constraints. The renormalization scale is fixed to $\mu = 4.2$ GeV. For comparison, we include the case with $(\mathcal{C}_7, \mathcal{C}_9, \mathcal{C}_{10})$ fixed at the SM values for which only nuisance parameters are varied (denoted by SM). The nuisance parameters are discarded when counting the degrees of freedom to compute the shown p -values [%] based on the statistics T_{like} and T_{pull} . [†] When comparing the posterior of the SM with the other modes, it has to be noted that the prior volume of $(\mathcal{C}_7, \mathcal{C}_9, \mathcal{C}_{10})$ is 6.68 in log units.

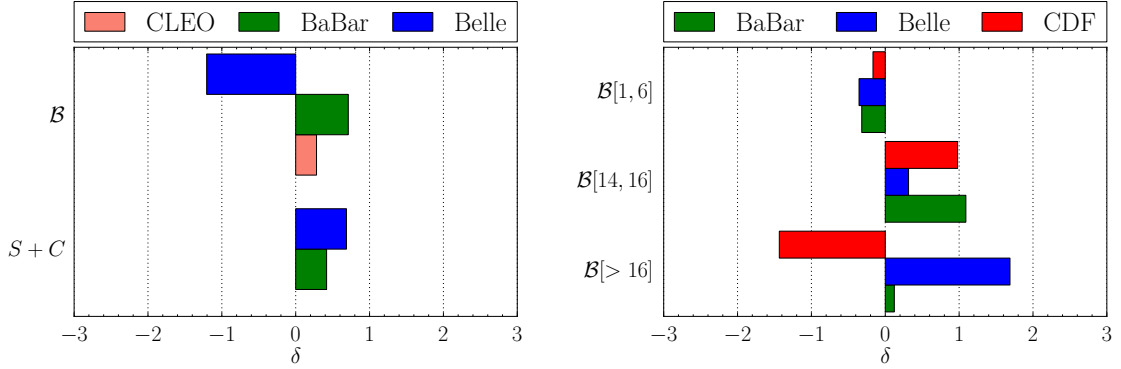


Figure 2. Pull values for observables in $B \rightarrow K^* \gamma$ [left] and $B \rightarrow K \bar{\ell} \ell$ [right] calculated at the best fit point. The pull definition for the correlated observables S and C permits only $\delta \geq 0$; for details see Sec. 3.5.

in favor of the simpler model. The effect persists if we cut the allowed range of each \mathcal{C}_i in half to exclude all but the SM-like solution. In conclusion, both the SM (with nuisance parameters allowed to vary) and our extension with real floating $\mathcal{C}_{7,9,10}$ fit the 59 experimental observations of rare B decays well. Since the extension does not provide any significant improvement, the simpler model should be preferred.

To study the dependence of our fit results on the priors, we use a second set of priors (*wide* priors). We scale the uncertainties of those parameters associated with form factors and unknown subleading contributions in Λ/m_b (Tab. 12) by a factor of three and adjust the parameter ranges accordingly. All other priors are kept the same. This choice includes the major sources of theory uncertainty and represents a pessimist's view of a) the validity of form factor results based on LCSR at low q^2 , b) their extrapolation to high q^2 values,

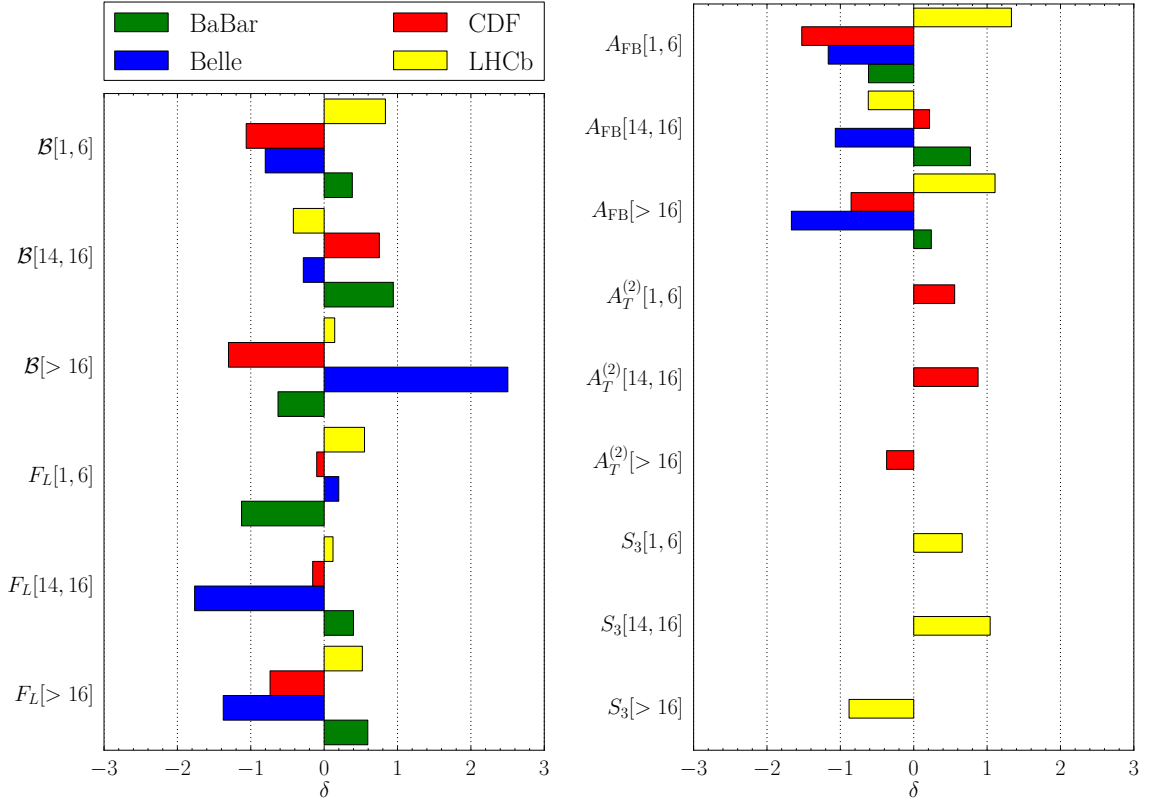


Figure 3. Pull values for observables in $B \rightarrow K^* \bar{\ell} \ell$ calculated at the best fit point.

	\mathcal{C}_7	\mathcal{C}_9	\mathcal{C}_{10}
68%	$[-0.34, -0.23] \cup [0.35, 0.45]$	$[-5.2, -4.0] \cup [3.1, 4.4]$	$[-4.4, -3.4] \cup [3.3, 4.3]$
95%	$[-0.41, -0.19] \cup [0.31, 0.52]$	$[-5.9, -3.5] \cup [2.6, 5.2]$	$[-4.8, -2.8] \cup [2.7, 4.7]$
max	$-0.28 \cup 0.40$	$-4.56 \cup 3.64$	$-3.92 \cup 3.86$
68%	$[-0.39, -0.19] \cup [0.30, 0.48]$	$[-5.6, -3.8] \cup [2.9, 5.1]$	$[-4.0, -2.5] \cup [2.6, 3.9]$
95%	$[-0.53, -0.13] \cup [0.24, 0.61]$	$[-6.7, -3.1] \cup [2.2, 6.2]$	$[-4.7, -1.9] \cup [2.0, 4.6]$
max	$-0.30 \cup 0.38$	$-4.64 \cup 3.84$	$-3.24 \cup 3.30$

Table 5. The 68% and 95% credibility intervals and the two local modes of the marginalized 1-dimensional posterior distributions of the Wilson coefficients $\mathcal{C}_{7,9,10}$ at $\mu = 4.2$ GeV for nominal [upper] and wide [lower] ranges of nuisance parameters (see App. B).

and c) subleading corrections exceeding expectations from power counting. The results of the fit at the low scale $\mu = 4.2$ GeV to all data with these new priors is shown in Fig. 4 alongside the corresponding 68%- and 95%-credibility regions of Fig. 1 for the two solutions in each of the three planes $\mathcal{C}_7 - \mathcal{C}_9$, $\mathcal{C}_7 - \mathcal{C}_{10}$ and $\mathcal{C}_9 - \mathcal{C}_{10}$. Most importantly, the fit

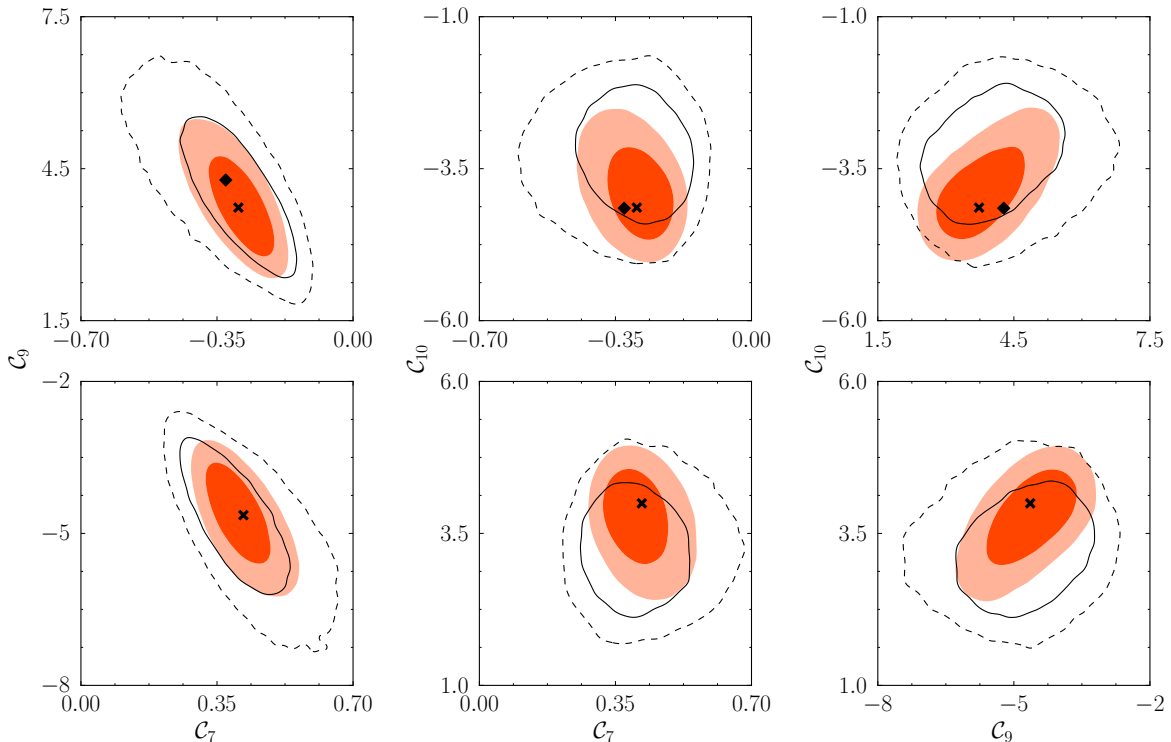


Figure 4. The marginalized 2-dimensional 68% (and 95%) credibility regions of the Wilson coefficients $\mathcal{C}_{7,9,10}$ at $\mu = 4.2$ GeV for the SM-like [top row] and sign-flipped solution [bottom row], arising from nominal ranges of Fig. 1 [red and light red, respectively] and wide ranges [solid and dashed contours, respectively] of the nuisance parameters. We indicate the values of $\mathcal{C}_{7,9,10}^{\text{SM}}$ in the SM [◆] and at the local maximum of the posterior [✕] resulting from nominal prior ranges in the respective region.

is stable and gives comparable results with both sets of priors thanks to the large number of experimental constraints. In all six planes, the area covered by the 68% region with wide priors is similar to that of the 95% region with nominal priors. While the two sets of regions are concentric in the $\mathcal{C}_7 - \mathcal{C}_9$ plane, there appears a rather hard cut-off at $|\mathcal{C}_{10}| \approx 5$ in the $\mathcal{C}_{7,9} - \mathcal{C}_{10}$ planes. For completeness, we list the set of smallest intervals and local maxima derived from the one-dimensional marginalized distributions for $\mathcal{C}_{7,9,10}$ for both sets of priors in Tab. 5. Our results for the 95% credibility intervals are compatible with those of Ref. [66]. More specifically, we find a larger interval for \mathcal{C}_7 , covering smaller values of $|\mathcal{C}_7|$. This is due to the use of $B \rightarrow X_s \gamma$ constraints that are used in Ref. [66], but not included in our work. However, with regard to $\mathcal{C}_{9,10}$, we find that our credibility intervals are 10–40% smaller. Compared to Ref. [66], we have added the 2012 results by LHCb and BaBar. The question arises if the inclusion of the inclusive decays $B \rightarrow X_s \gamma$ and $B \rightarrow X_s \bar{\ell} \ell$ could further shrink the $\mathcal{C}_{9,10}$ credibility intervals.

From the allowed ranges for $\mathcal{C}_{7,9,10}$, we can estimate limits on the scale of generic

	Λ_7^{NP} [TeV]	Λ_9^{NP} [TeV]	Λ_{10}^{NP} [TeV]
SM-like	29, 38	28, 37	30, 44
SM-sign-flipped	12, 13	11, 13	12, 13

Table 6. Constraints on the NP scale Λ_i^{NP} ($i = 7, 9, 10$) assuming generic flavor violation at tree level using the 95% credibility region from Tab. 5. Several possibilities arise from destructive and constructive interference of the SM with SM-like and SM-sign-flipped solutions.

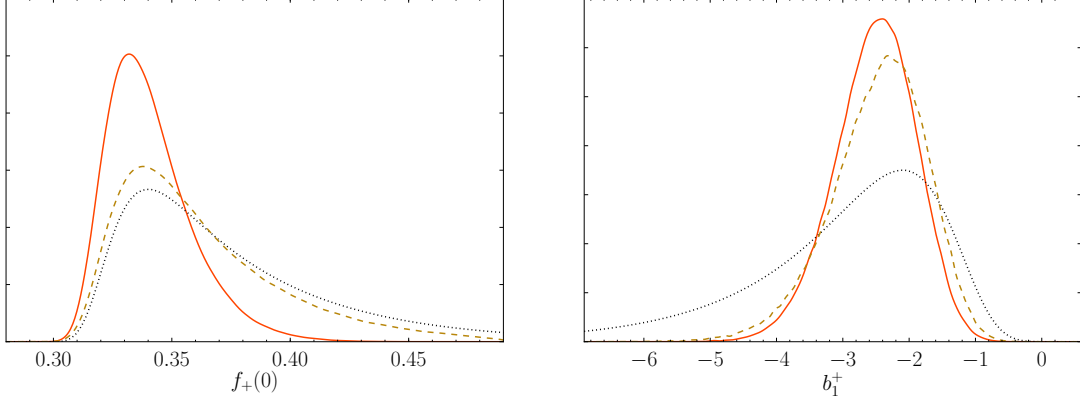


Figure 5. Prior [dotted] and posterior distributions of the nuisance parameters $f_+(0)$ [left] and b_+^1 [right], governing the normalization and the q^2 shape of the $B \rightarrow K$ form factor $f_+(q^2)$, respectively. We show the posterior using $B \rightarrow K \bar{\ell} \ell$ data only [dashed] vs all data [solid].

flavor-changing neutral currents at tree level, described by

$$\mathcal{H}_{\text{eff}} = \sum_{i=7,9,10} \frac{\tilde{\mathcal{O}}_i}{(\Lambda_i^{\text{NP}})^2}, \quad (4.3)$$

$$\tilde{\mathcal{O}}_7 = m_b [\bar{s} \sigma_{\mu\nu} P_R b] F^{\mu\nu}, \quad \tilde{\mathcal{O}}_{9,10} = [\bar{s} \gamma_\mu P_L b] [\bar{\ell} \gamma^\mu (1, \gamma_5) \ell]. \quad (4.4)$$

Using $\mathcal{C}_i = \mathcal{C}_i^{\text{SM}} + \mathcal{C}_i^{\text{NP}}$ and setting \mathcal{C}_i to the boundary values of the 95% intervals (nominal priors), we extract $\mathcal{C}_i^{\text{NP}}$. By matching (4.3) with (2.1) and (2.2), we extract the minimum scale Λ_i^{NP} for both constructive and destructive interference with the SM; see Tab. 6. The resulting scales above which NP “is still allowed” are similar to those found in previous analyses [64] and [66].

So far, we discussed the fit results for the Wilson coefficients $\mathcal{C}_{7,9,10}$ that enter most, but not all of the observables. Exceptions are those of $B \rightarrow K^* \gamma$, which depends only on \mathcal{C}_7 , and $B_s \rightarrow \bar{\mu} \mu$, which depends only on \mathcal{C}_{10} . The marginalized distributions in the $\mathcal{C}_9 - \mathcal{C}_{10}$ plane of Fig. 1 show that, compared to $B \rightarrow K^*$, the fit with $B \rightarrow K$ only measurements prefers a smaller value of $|\mathcal{C}_9|^2 + |\mathcal{C}_{10}|^2$; the marginal modes (not shown) are near $\mathcal{C}_9 = 0$, $\mathcal{C}_{10} = \pm 5$. Since the $B \rightarrow K^*$ constraints dominate the combination, a “tension” arises.

Let us now discuss the role that the nuisance parameters play in the fit. First, we note that the posterior distributions of the common nuisance parameters — those that are not

specific to rare $b \rightarrow s$ decays, like the CKM parameters and the c and b quark $\overline{\text{MS}}$ masses — do not deviate from their prior distributions given in Tab. 10. This is mainly due to the strong prior knowledge from other measurements and the comparatively low precision of both experimental and other, mostly hadronic, theory inputs in the rare $b \rightarrow s$ decays.

Second, we consider the remaining hadronic nuisance parameters of form factors and subleading corrections, for which the priors are based mostly on educated guesses rather than precise knowledge. Because $B \rightarrow K$ and $B \rightarrow K^*$ form factors enter observables at both low and high q^2 , they are determined by all the $B \rightarrow K \bar{\ell}\ell$ and $B \rightarrow K^*(\gamma, \bar{\ell}\ell)$ observables respectively. In contrast, the parametrization of unknown subleading Λ/m_b corrections is different at low and high q^2 (and naturally in $B \rightarrow K$ and $B \rightarrow K^*$ decays). Since subleading corrections at high q^2 receive further parametric suppression by either $\mathcal{C}_7/\mathcal{C}_9$ or α_s [44, 57], the corresponding observables at high q^2 are rather weakly dependent on them. In contrast, at low q^2 large effects are not surprising.

Therefore, we expect a significant update to knowledge of form factors to accommodate the tension between $B \rightarrow K$ and $B \rightarrow K^*$ constraints. Any remaining tension should be visible in low- q^2 subleading corrections.

Let us first consider the posterior distributions of the two nuisance parameters $f_+(0)$ and b_1^+ , which enter the q^2 parametrization of the $B \rightarrow K$ form factor $f_+(q^2)$ (see (B.3) and priors in Tab. 12 from LCSR results [41]). The q^2 shape of the form factor is controlled by b_1^+ . The low- and high- q^2 data of the $B \rightarrow K \bar{\ell}\ell$ branching fraction (Tab. 2) give rise to a narrower posterior compared to the prior distribution in Fig. 5, which does not change much when using only $B \rightarrow K \bar{\ell}\ell$ data or combining it with $B \rightarrow K^* \bar{\ell}\ell$. This preference also appears when choosing the wide set of ranges for the prior distributions of the nuisance parameters, demonstrating that the data suppress the tails in the prior of b_1^+ . Concerning $f_+(0)$, which corresponds to the normalization of the form factor, we observe a strong preference for low values in the posterior distribution in Fig. 5. However, this preference almost disappears when only $B \rightarrow K \bar{\ell}\ell$ data is used in the fit. This behavior persists even when allowing for wider prior ranges, and is easily understood in terms of the above-mentioned tension.

We also find strong modifications of the posterior with respect to prior distributions for the three scale factors $\zeta_{A_1, A_2, V}$ entering the form factors A_1, A_2, V in $B \rightarrow K^*$. The posteriors are shown in Fig. 6 along with the common prior distribution. Of the three, A_1 is known most accurately after the fit, while A_2 and V are simultaneously shifted and compressed. Using all constraints, A_1, A_2, V are shifted towards higher values, but without $B \rightarrow K$ constraints, the shift actually points in the opposite direction. Again, the positive shift serves to reduce the tension and allows a good fit to all constraints with values of $\mathcal{C}_{9,10}$ smaller than required by the $B \rightarrow K^*$ constraints alone.

The parameters describing subleading phases are mostly unaffected by the fit. All phases come out with a flat distribution, indicating that they could have been omitted from the fit without any consequences.

The largest update to knowledge of subleading parameters occurs for the scale factor of the transversity amplitudes $A_{0,\perp}^L$ (B.3) describing the $B \rightarrow K^*$ decays, with a downward shift of about 10% and a slight reduction of variance. We observe this effect only in

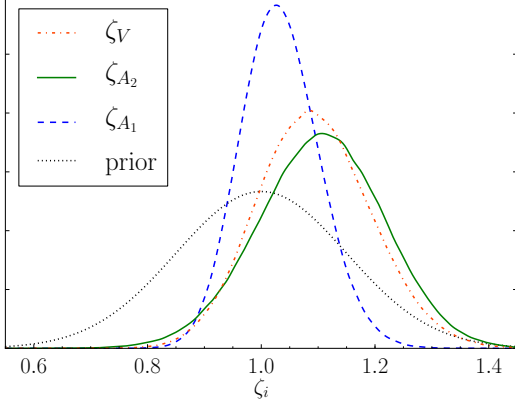


Figure 6. Posterior distributions of the fit with all data for the nuisance parameters $\zeta_{A_1, A_2, V}$ serving as scale factors to the corresponding $B \rightarrow K^*$ form factors. The common prior is indicated.

q^2 -bin	$\langle A_T^{(3)} \rangle$	$\langle A_T^{(4)} \rangle$	$\langle A_T^{(5)} \rangle$
[1.0, 6.0]	$0.454^{+0.081}_{-0.086} {}^{+0.181}_{-0.158}$	$0.565^{+0.156}_{-0.121} {}^{+0.355}_{-0.234}$	$0.468^{+0.019}_{-0.025} {}^{+0.030}_{-0.056}$
q^2 -bin	$\langle A_T^{(\text{re})} \rangle$	$\langle H_T^{(1)} \rangle$	$\langle H_T^{(2)} \rangle$
[1.0, 6.0]	$0.33^{+0.14}_{-0.10} {}^{+0.25}_{-0.22}$	$0.441^{+0.055}_{-0.058} {}^{+0.105}_{-0.113}$	$-0.271^{+0.057}_{-0.060} {}^{+0.117}_{-0.117}$

Table 7. Predictions of unmeasured, optimized observables based on global fit output integrated over the large recoil region. We list the most probable value, the smallest 68% and 95% intervals.

the fit with all observables. Neither A_i^R nor $B \rightarrow K$ subleading parameters are updated significantly in any of the fits. A_i^R has little effect compared to A_i^L because the observables depend on $A_i^{L,R} \propto \mathcal{C}_9 \mp \mathcal{C}_{10}$, and $\mathcal{C}_9 \approx -\mathcal{C}_{10}$.

In summary, we do not observe a drastic update of any nuisance parameter, showing that the fit is stable³. The uncertainty on the form factors and some subleading corrections is reduced by the data, but the most likely values are shifted due to the tension between $B \rightarrow K$ and $B \rightarrow K^*$ constraints. More theory input is required to reduce the uncertainty on the remaining subleading corrections.

4.2 Predictions

As outlined in Sec. 2.2.3, the angular distribution of $B \rightarrow K^*(\rightarrow K\pi)\bar{\ell}\ell$ allows one to form optimized observables, which have reduced form factor uncertainties and may exhibit sensitivity to a particular type of new physics. Currently, no measurements of these observables are available. We provide predictions at low and high q^2 within the scenario of the SM operator basis, taking into account the present data. Consequently, future observations outside the predicted ranges would indicate physics beyond the considered scenario.

The predictions of $A_T^{(3,4,5,\text{re})}$ and $H_T^{(1,2)}$ at low q^2 are given in q^2 -integrated form for the bin $q^2 \in [1, 6]$ GeV² in Tab. 7. In addition, Fig. 7 shows the results of the 5 sub-bins with a bin width of 1 GeV², as used in the first measurement of the lepton A_{FB} of $B \rightarrow K^*\bar{\ell}\ell$ by LHCb [18]. The observables $A_T^{(3,4)}$ have been chosen due to their sensitivity

³For the suppressed solutions, scale factors for $B \rightarrow K^*$ form factors and A_{\perp}^L shift by $\mathcal{O}(15\%)$ and A_{\parallel}^L even peaks at the left boundary.

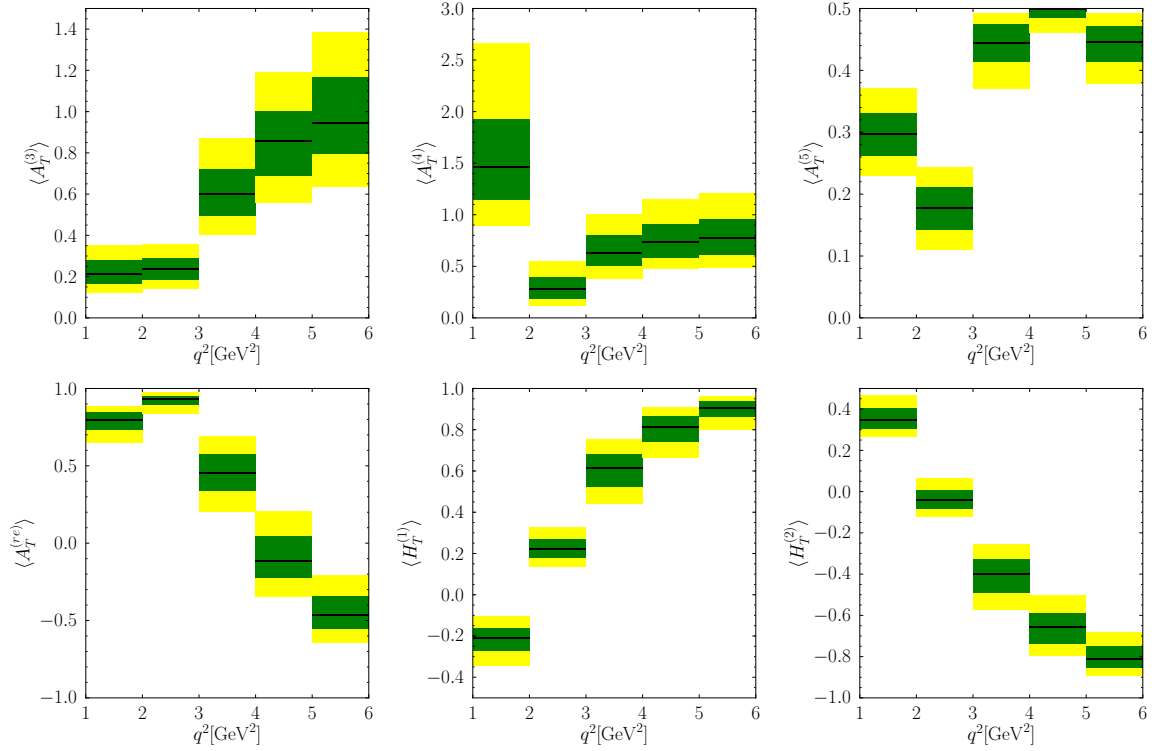


Figure 7. Predictions of unmeasured optimized observables at large recoil based on the global fit output. We show the most probable value [solid black line] as well as the smallest 68% (green) and 95% (yellow) intervals of the q^2 -integrated observables.

q^2 -bin	$\langle H_T^{(1)} \rangle$	$\langle H_T^{(2)} \rangle$	$\langle H_T^{(3)} \rangle$
[14.18, 16]	0.99969 $^{+0.00009}_{-0.00011}$ $^{+0.00015}_{-0.00026}$	-0.9843 $^{+0.0023}_{-0.0022}$ $^{+0.0056}_{-0.0039}$	-0.9837 $^{+0.0022}_{-0.0019}$ $^{+0.0053}_{-0.0033}$
[16, 19.21]	0.99896 $^{+0.00025}_{-0.00032}$ $^{+0.00044}_{-0.00076}$	-0.9704 $^{+0.0018}_{-0.0019}$ $^{+0.0042}_{-0.0037}$	-0.9614 $^{+0.0015}_{-0.0012}$ $^{+0.0037}_{-0.0021}$
[14.18, 19.21]	0.99772 $^{+0.00058}_{-0.00078}$ $^{+0.00105}_{-0.00179}$	-0.9733 $^{+0.0027}_{-0.0023}$ $^{+0.0057}_{-0.0043}$	-0.9608 $^{+0.0019}_{-0.0015}$ $^{+0.0045}_{-0.0027}$

Table 8. Predictions of unmeasured, optimized observables based on global fit output for the two conventional bins and the entire low recoil region. We list the most probable value, the smallest 68% and 95% intervals.

to the chirality-flipped \mathcal{C}'_7 [53]. The large discontinuity of $A_T^{(4)}$ in $q^2 \in [1, 3]$ GeV² is caused by the zero crossing of J_4 in its denominator (2.20). The observable $A_T^{(5)}$ is restricted by construction to take values in $[-0.5, 0.5]$ and reaches its maximal value at the zero crossing of the lepton A_{FB} in the bin $q^2 \in [4, 5]$ GeV² [56]. Its shape is sensitive to new physics contributions of the Wilson coefficients. Note that the theory uncertainty is at a minimum when $A_T^{(5)}$ approaches 0.5.

The observable $A_T^{(\text{re})}$ reaches its maximal value of about 1.0 in $q^2 \in [2, 3]$ GeV² and has the very same zero crossing as the leptonic A_{FB} . Our results are in qualitative agreement

with [60], who stressed that the deviation of the maximal value from 1.0 and its position are sensitive to new physics. The observables $H_T^{(1,2)}$ were first proposed for the high- q^2 region [57] as long-distance free observables. In addition, $H_T^{(1)}$ is also short-distance free, with $|H_T^{(1)}(q^2)| = 1$, depending only on the sign of a form factor. Recently it was shown that at low q^2 , form factors also cancel in $H_T^{(1,2)}$ [61]. Each has a zero crossing in the region $q^2 \in [1, 3] \text{ GeV}^2$ that is the very same as in the CP-averaged normalized observables J_4/Γ and J_5/Γ [54, 55]. For $H_T^{(1)}$, one observes the rise towards ≈ 1.0 for rising q^2 .

At high- q^2 , the situation is more restrictive, and within the scenario of the SM operator basis, there are only three optimized observables $H_T^{(1,2,3)}$ [57]. The predictions for three q^2 bins are given in Tab. 8. Besides $|H_T^{(1)}(q^2)| = 1$, we have the additional relation $H_T^{(2)}(q^2) = H_T^{(3)}(q^2)$. Small deviations in the predictions of $\langle H_T^{(1,2,3)} \rangle$ arise from separate q^2 -integration of J_i (see (2.7) and below), such that the equality does not hold exactly. Any large experimental deviation from the prediction $|H_T^{(1)}(q^2)| = 1$ would signal a breakdown of the OPE [101]. The observables $H_T^{(2,3)}(q^2)$ are given by the short-distance ratio [57]

$$H_T^{(2,3)}(q^2) = \frac{2 \operatorname{Re} [\mathcal{C}_{79}^{\text{eff}}(q^2) \mathcal{C}_{10}^*]}{|\mathcal{C}_{79}^{\text{eff}}(q^2)|^2 + |\mathcal{C}_{10}|^2} = \cos(\varphi_{79}(q^2) - \varphi_{10}) \frac{2r}{1+r^2} \quad (4.5)$$

with

$$\mathcal{C}_{79}^{\text{eff}}(q^2) = \mathcal{C}_9^{\text{eff}}(q^2) + \kappa \frac{2m_b^2}{q^2} \mathcal{C}_7^{\text{eff}}(q^2), \quad r(q^2) = \frac{|\mathcal{C}_{79}^{\text{eff}}(q^2)|}{|\mathcal{C}_{10}|} \quad (4.6)$$

and $\mathcal{C}_i^{\text{eff}}(q^2)$ and the factor $\kappa = 1 + \mathcal{O}(\alpha_s)$ of the improved Isgur-Wise form factor relation defined in [57]. In the SM, $\mathcal{C}_{10}^{\text{SM}} \approx -4.2$ and therefore its phase $\varphi_{10} = \pi$. The q^2 dependence of the sum of the effective Wilson coefficients $\mathcal{C}_{79}^{\text{eff}}(q^2)$ is rather weak and its imaginary parts at NLO in QCD small [59], such that $\varphi_{79}(q^2) \approx 0$; whereas the magnitudes of the Wilson coefficients are $\mathcal{C}_9^{\text{SM}} \approx +4.2$ and $\mathcal{C}_7^{\text{SM}} \approx -0.3$, and lead to $r \approx 1$ and $\cos(\varphi_{79}(q^2) - \varphi_{10}) \approx -1$. Therefore, $H_T^{(2,3)}$ test roughly the ratio of $|\mathcal{C}_9|/|\mathcal{C}_{10}|$ within our scenario of the SM operator basis and real Wilson coefficients. The results in Tab. 8 show that current data do not allow for deviations from the SM prediction. We remark that the prediction of $\langle H_T^{(1)} \rangle$ is based on the OPE and is expected to be 1 at any particular value of q^2 . Therefore, our results just reflect how precisely the form factor and the modeled subleading corrections cancel for the q^2 -integrated version when taking into account the update of our knowledge of the nuisance parameters due to the experimental information.

Although SM predictions have been given previously [54, 64, 66], our Monte Carlo approach described in Sec. 3.4 provides several improvements with respect to the standard procedure to estimate theory uncertainties. Usually observables $X(\nu)$ are computed at three values of a single parameter ν : at the central value ν_{cen} and at $(\nu_{\text{cen}})_{-a}^{+b}$. The changes in the predictions of X are then interpreted as the associated uncertainty: $\sigma_{+,-} = |X(\nu_{\text{cen}}) - X(\nu_{\text{cen}})_{-a}^{+b}|$, and the central value of X is assumed to be $X(\nu_{\text{cen}})$. In the presence of several parameters, the respective uncertainties are then combined either linearly or in quadrature into a total uncertainty. In contrast to this so-called min-max approach, we vary all parameters at the same time and thus automatically take correlations into account. Our

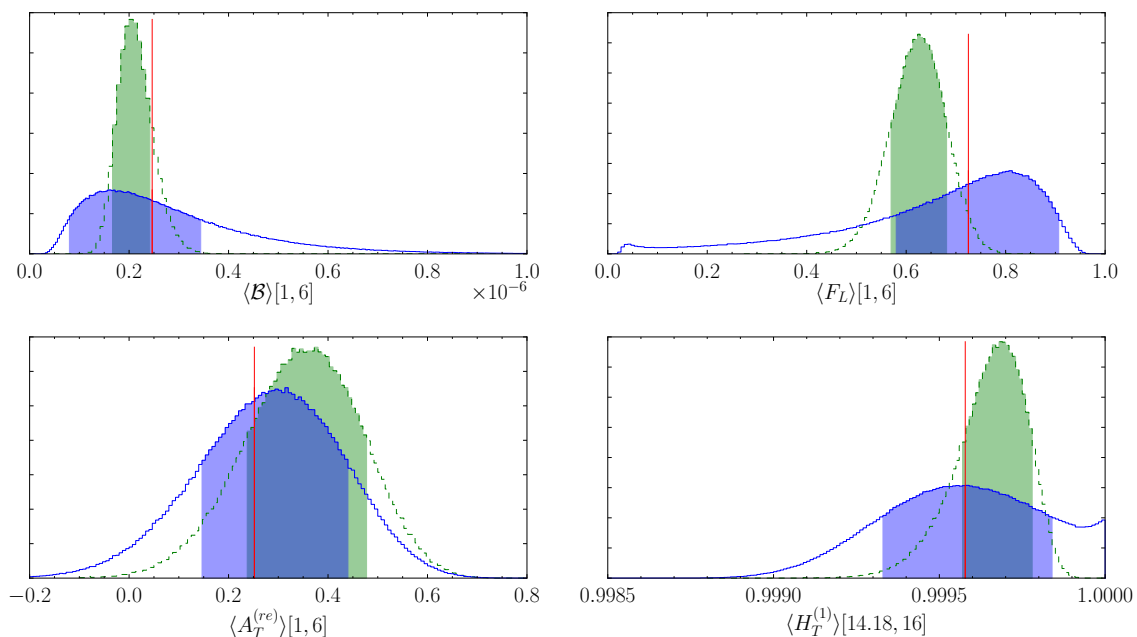


Figure 8. Probability distributions of the SM predictions of q^2 -integrated observables in the $B \rightarrow K^* \bar{\ell} \ell$ decay, when varying nuisance parameters within their allowed prior ranges [solid, blue]. The shaded region is the 68% interval and the vertical (red) line indicates the prediction when using central values of nuisance parameters. Also shown are the predictions based on posterior distributions [dashed, green] determined by the experimental data, allowing also for NP in $\mathcal{C}_{7,9,10}$ in the fit.

intervals have a strict probabilistic interpretation as Bayesian credibility intervals, and the procedure automatically takes care of non-linearities. As a simple example consider the quadratic dependence of a branching ratio $\mathcal{B} \propto f^2$ on a decay constant or form factor f . Assuming a Gaussian prior distribution of f , $p(\mathcal{B})$ is the (asymmetric) χ^2 -distribution with one degree of freedom. Typical examples of asymmetry can be seen in in Fig. 8 for $\langle \mathcal{B} \rangle[1, 6]$ and $\langle F_L \rangle[1, 6]$ (blue, solid) of the decay $B \rightarrow K^* \bar{\ell} \ell$, where the maximum of the distribution deviates from the vertical (red) line that indicates the prediction obtained by using central values for all nuisance parameters; i.e., the position of the maxima of their priors. This behavior is not present in $\langle A_T^{(re)} \rangle[1, 6]$ since there, form factors cancel; likewise in $\langle H_T^{(1)} \rangle[14.18, 16]$. We list the modes and 68% intervals for a number of observables in App. C in Tab. 13 and Tab. 14, but stress that the uncertainty of an observable X is best described by the probability distribution $p(X)$. In the simplest case, $p(X)$ can be described by the 68% interval and the mode, but in general, it contains much more information as demonstrated in Fig. 8.

Let us finally compare the SM predictions of observables based on the prior information with predictions based on posterior distributions as determined by experimental data and allowing also for NP in $\mathcal{C}_{7,9,10}$. Our posterior findings are overlaid on the SM predictions for the examples in Fig. 8. Although NP contributions to the Wilson coefficients are included, in all cases the posterior distributions are narrower than the SM prediction based

on prior knowledge only. Obviously, the additional information from data on the nuisance parameters updates our knowledge on quantities $\langle \mathcal{B} \rangle[1,6]$ and $\langle F_L \rangle[1,6]$, which served as inputs to the fit.

As described in the previous section, both fit solutions for $\mathcal{C}_{7,9,10}$ give an overall good description of the experimental data. For the optimized observable $\langle A_T^{(\text{re})} \rangle[1,6]$, they yield a prediction of similar range, shifted slightly towards larger values, compared to the SM prediction based on prior knowledge alone. The same situation emerges for the other optimized observables, which are free of form factor uncertainties — compare Tab. 7 and 13 for low- q^2 as well as Tab. 8 and 14 for high q^2 — and main uncertainties are due to lacking subleading corrections. At this stage, better prior knowledge on the nuisance parameters is needed. This will help to distinguish new physics from the SM with the help of optimized observables, in the scenario of the SM operator basis with real Wilson coefficients. However, any experimental observation outside of the predicted range would point strongly to an extended scenario.

5 Conclusion

We perform a fit of the short-distance couplings $\mathcal{C}_{7,9,10}$ appearing in the effective theory of $\Delta B=1$ decays describing $b \rightarrow s\gamma$ and $b \rightarrow s\bar{\ell}\ell$ transitions, assuming $\mathcal{C}_{7,9,10}$ to be real valued. For the first time, we include all relevant theory uncertainties in the analysis by means of nuisance parameters. Measurements of exclusive rare decays $B \rightarrow K^*\gamma$, $B \rightarrow K^{(*)}\bar{\ell}\ell$ and $B_s \rightarrow \bar{\mu}\mu$ obtained by CLEO, BaBar, Belle, CDF, and LHCb serve as experimental inputs. Besides presenting credibility intervals for the Wilson coefficients, we analyze the goodness of fit of the obtained solutions. For the best-fit solution, we show the pull values for all measurements in Figures 2 and 3. We use a novel combination of Markov Chain Monte Carlo and adaptive importance sampling methods in order to cope with the high dimensionality of the parameter space (~ 30) and the multimodal posterior distribution. With this approach, we can massively parallelize the costly evaluation of the posterior. Our results should simplify subsequent model-dependent studies; we are happy to provide the fit output in a suitable format upon request.

The credibility intervals of the marginalized one- and two-dimensional posterior distributions of $\mathcal{C}_{7,9,10}$ are the main results of our fit, given in Tab. 5 and Fig. 1. Due to a discrete symmetry, a SM-like and a flipped-sign solution remain with posterior mass ratio of roughly 51% over 49%. Other local maxima exist, but their posterior masses are negligible. The SM values $\mathcal{C}_{7,9,10}^{\text{SM}}$ are close to the best-fit point. Both solutions as well as the SM itself provide a good fit of the data. Judging by the Bayes factor as model comparison criterion, the data clearly favor the plain SM over a model with arbitrary real $\mathcal{C}_{7,9,10}$ — a tribute to Occam’s razor. Thus, from a purely statistical point of view, even the simplest model-independent extension of the SM is not necessary to describe the current data. We emphasize that the presence of the sign-flipped solution still allows large NP contributions to the Wilson coefficients. However, the degeneracy of the observables does not allow us to distinguish them easily. This degeneracy is mildly broken by contributions of 4-quark operators, typically included in the effective Wilson coefficients $\mathcal{C}_{7,9} \rightarrow \mathcal{C}_{7,9}^{\text{eff}}$. Assuming im-

proved theory uncertainties and current experimental central values in $B \rightarrow K^* \gamma$, the fit suggests that the additional information on $\mathcal{C}_7^{\text{eff}}$ enhances the SM-like solution over the flipped-sign solution. We expect a reduced theory uncertainty when including $B \rightarrow X_s \gamma$.

We provide updated predictions within the SM of selected observables in the angular distribution of $B \rightarrow K^*(\rightarrow K\pi) \bar{\ell}\ell$. Based on prior knowledge only, we obtain reduced theory uncertainties due the improved handling of uncertainty propagation, observing that the central values of previous analyses [57, 59, 66] are contained in the smallest 68% regions.

Based on the fit output, we predict ranges for currently unmeasured observables that exhibit a reduced form factor dependence. Surprisingly, the predictions based on the fit output yield smaller ranges than SM predictions based on prior knowledge. The extra variance due to Wilson coefficients is more than compensated for by the reduced uncertainties as the fit constrains some of the nuisance parameters and yields the correlation between all parameters.

We observe that a fit with current $B \rightarrow K \bar{\ell}\ell$ constraints prefers smaller values of $\mathcal{C}_{9,10}$ than a fit with the $B \rightarrow K^* \bar{\ell}\ell$ constraints. Including both sets of constraints, the fit accommodates this tension by shifting the $B \rightarrow K$ form factors towards smaller values, and the $B \rightarrow K^*$ form factors towards larger values.

Future analyses can improve the fit by including results for the inclusive decays $B \rightarrow X_s \gamma$ as well as $B \rightarrow X_s \bar{\ell}\ell$. Besides the inclusion of additional observables, further enhancements could arise when using an alternative parametrization of $B \rightarrow K^*$ form factors; cf [114].

Acknowledgments

We would like to thank our colleagues from LHCb Diego Martinez Santos and Johannes Albrecht for providing help on the $B_s \rightarrow \bar{\mu}\mu$ results, as well as Ulrik Egede, Thomas Blake, Mitesh Patel, and Will Reece for valuable comments on the data of $B \rightarrow K^* \bar{\ell}\ell$. We are grateful to Martin Kilbinger for his support on questions related to Population Monte Carlo and “pmclib”. We thank Gudrun Hiller and Allen Caldwell for many stimulating discussions and remarks as well as their advice on the draft. Daniel Greenwald provided innumerable corrections to enhance legibility and clarity of this article. Finally, we thank Martin Gorbahn for his support of our work.

A Numerical Input

The numerical input values of parameters are listed in Tab. 9, for which the uncertainties have not been included since they are either very small or they enter in numerically subleading contributions to the observables of interest.

The theory predictions of all the relevant semileptonic and radiative processes at large recoil are based on the QCDF results [38, 40]. These include the usage of the Light Cone Distribution Amplitudes (LCDA) of the involved kaons which are parametrized in terms

$\alpha_s(M_Z)$	0.11762		m_μ	0.106 GeV	[91]
$\alpha_e(m_b)$	1/133		m_t^{pole}	173.3 GeV	[102]
$\sin^2\theta_W$	0.23116	[91]	M_W	80.399 GeV	[91]
τ_{B^+}	1.638 ps	[91]	τ_{B^0}	1.525 ps	[91]
M_{B^+}	5.2792 GeV	[91]	M_{B^0}	5.2795 GeV	[91]
M_{K^+}	0.494 GeV	[91]	M_{K^0}	0.498 GeV	[91]
$M_{K^{*+}}$	0.892 GeV	[91]	$M_{K^{*0}}$	0.896 GeV	[91]
τ_{B_s}	1.472 ps	[91]	M_{B_s}	5.3663 GeV	[91]
$\lambda_{B,+}$	0.485 GeV		$f_{B^{0,+}}$	0.212 GeV	[103]
f_K	0.1561 GeV				
$f_{K_\perp^*}(2 \text{ GeV})$	0.173 GeV		$f_{K_\parallel^*}$	0.217 GeV	
$a_1(K)$	0.048		$a_2(K)$	0.174	
$a_1(K_\perp^*)$	0.1		$a_2(K_\perp^*)$	0.1	
$a_1(K_\parallel^*)$	0.1		$a_2(K_\parallel^*)$	0.1	

Table 9. The numerical input used in the analysis. The mass of the strange quark has been neglected throughout. τ_{B^0} (τ_{B^+}) denotes the lifetime of the neutral (charged) B meson. The following parameters appear in expressions of $B \rightarrow (K, K^*) \bar{\ell}\ell$ at large recoil: $\lambda_{B,+}$ denotes the first inverse moment of the B -meson distribution amplitude, whereas f_M the decay constants and $a_{1,2}(M)$ are the first two Gegenbauer moments of the LCDA's of the respective Kaon states $M = K, K_\perp^*, K_\parallel^*$.

of Gegenbauer moments $a_n(M)$ ($M = K, K_\perp^*, K_\parallel^*$). In this work, we include terms in the expansion in Gegenbauer moments up to $n = 2$, using the central values in Tab. 9.

Since the $a_n(M)$ also enter the computation of the $B \rightarrow K^*$ form factors via LC sum rules [107], variation of the former would lead to double counting. Furthermore, the residual influence of the $a_n(M)$ on the observables is small compared to that of other parameters. We therefore do not vary the Gegenbauer moments.

In addition, QCDF makes use of the decay constants f_M ($M = K, K_\perp^*, K_\parallel^*$), which enter in numerically subleading contributions. The central values are listed in Tab. 9.

B Nuisance Parameters

In this section we present the nuisance parameters that are considered in this work and contribute the main uncertainties in theory predictions. All the priors of these parameters are clipped to the parameter range that corresponds to their respective 3σ interval. For the sake of readability, we categorize the individual nuisance parameters.

B.1 Common Nuisance Parameters

The common nuisance parameters are those that enter most of the observables and are not specific for rare $b \rightarrow s$ decays. These are the elements of the Cabibbo-Kobayashi-Maskawa (CKM) quark-mixing matrix and the b and c quark masses.

A	0.804 ± 0.010	[92]	λ	0.22535 ± 0.00065	[92]
$\bar{\rho}$	0.111 ± 0.070	[92]	$\bar{\eta}$	0.381 ± 0.030	[92]
$m_c(m_c)$	$(1.27^{+0.07}_{-0.09})$ GeV	[91]	$m_b(m_b)$	$(4.19^{+0.18}_{-0.06})$ GeV	[91]

Table 10. Common nuisance parameters: The CKM Wolfenstein parameter values as obtained from the CKM *tree-level fit*, cf. Sec. 3.2.

For the purpose of the fit of rare $b \rightarrow s$ decays, we take the CKM matrix elements from other observables such as tree decays. We parametrize the CKM matrix elements using the Wolfenstein parametrization to $\mathcal{O}(\lambda^9)$ [108] and use the results of the tree-level fit of the UTfit collaboration [92] as priors in the fit of $b \rightarrow s$ decays. In this way, we include non-SM effects, but assume they do not affect tree-level decays. However, we use the results of the SM CKM fit in order to determine the uncertainties of observables in the framework of the SM in Sec. 4.2. Note that the CKM matrix elements only enter the branching ratios of $B \rightarrow K^{(*)} + (\gamma, \bar{\ell}\ell)$ decays in the combination $V_{tb}V_{ts}^*$. Although numerically negligible, the combination $V_{ub}V_{us}^*$ entering all observables is included in the analysis. It becomes relevant only for CP-asymmetric observables. All priors are Gaussian, with their 1σ ranges given in Tab. 10.

The values of the quark masses m_b and m_c enter most observables. In order to account for the asymmetric errors, we use LogGamma distributions (see Sec. D.2) as priors whose modes and 68%-probability intervals match the values given in Tab. 10.

B.2 $B \rightarrow K^{(*)}$ Form Factors and f_{B_s}

The heavy-to-light form factors $f_{+,T,0}$ for $B \rightarrow K$ and $V, A_{0,1,2}$, and $T_{1,2,3}$ for $B \rightarrow K^*$ transitions present a major source of uncertainty in predictions of rare exclusive B decays. They are functions of the dilepton invariant mass q^2 and we adopt the definition used in [40, 41, 48, 107]. Due to the application of form factor relations at large and low recoil, only f_+ enters $B \rightarrow K$ and V and $A_{1,2}$ enter $B \rightarrow K^*$ transitions⁴. The application of form factor relations introduces uncertainties of order Λ_{QCD}/m_b that will be discussed in App. B.3.

Currently, the form factors are only known from Light Cone Sum Rules (LCSR) which are applicable at low q^2 . Lattice QCD can provide results at high q^2 , where quenched results for some form factors [109, 110] are available and some preliminary unquenched results have been reported in [111–113]. An extensive discussion of the q^2 -shape parametrization using series expansion and a fit to low- q^2 LCSR combined with high- q^2 lattice results (when available) can be found in [114].

With regard to $B \rightarrow K^*$ form factors $V, A_{1,2}$, we use the LCSR results at low- q^2 as given in [107], where the extrapolation to high- q^2 is based on a (multi-)pole ansatz

$$V = \frac{r_1}{1 - q^2/m_R^2} + \frac{r_2}{1 - q^2/m_{\text{fit}}^2}, \quad (\text{B.1})$$

⁴The form factors f_0 and A_0 do not contribute within the framework of the SM operator basis, up to negligible terms suppressed by m_ℓ^2/q^2 .

	r_1	r_2	m_R^2 [GeV ²]	m_{fit}^2 [GeV ²]
V	0.923	-0.511	5.32 ²	49.40
A_1	—	0.290	—	40.38
A_2	-0.084	0.343	—	52.00

Table 11. The parameters of the form factors V and $A_{1,2}$.

$$A_1 = \frac{r_2}{1 - q^2/m_{\text{fit}}^2}, \quad A_2 = \frac{r_1}{1 - q^2/m_{\text{fit}}^2} + \frac{r_2}{(1 - q^2/m_{\text{fit}}^2)^2},$$

and the numerical values of the parameters given in Tab. 11. We do not vary these parameters themselves as they strongly depend on the LCSR analysis, but rather assign one multiplicative scaling factor ζ_i per form factor ($i = V, A_1, A_2$) to model the respective uncertainty such that the value $\zeta_i = 1.0$ corresponds to the central value of the form factor. A Gaussian prior is assigned to these nuisance parameters, which has a width of $\sigma = 0.15$ (i.e., 15% uncertainty) and its support extends up to 3σ (i.e., maximally 45% uncertainty), outside of which the prior is set to zero (see Tab. 12). Note that in this way we do not vary the q^2 shape of the form factors. At large recoil, two universal form factors [40] appear

$$\xi_{\perp} \equiv \frac{M_B}{M_B + M_{K^*}} V, \quad \xi_{\parallel} \equiv \frac{M_B + M_{K^*}}{2E_{K^*}} A_1 - \frac{M_B - M_{K^*}}{M_B} A_2, \quad (\text{B.2})$$

and their variation is obtained by the uncorrelated variation of V and $A_{1,2}$ as described above.

Since we calculate the $B \rightarrow K^* \gamma$ matrix element within QCDF for $q^2 = 0$, all nuisance parameters that affect the process $B \rightarrow K^* \bar{\ell} \ell$ in the large recoil region likewise affect the radiative process, as far as they are applicable.

With regard to the $B \rightarrow K$ form factor f_+ , we use the BCL parametrization [115] of the LCSR results [41]

$$f_+(q^2) = \frac{f_+(0)}{1 - q^2/M_{\text{res},+}^2} \left[1 + b_1^+ \left(z(q^2) - z(0) + \frac{1}{2} [z(q^2)^2 - z(0)^2] \right) \right], \quad (\text{B.3})$$

$$z(s) = \frac{\sqrt{\tau_+ - s} - \sqrt{\tau_+ - \tau_0}}{\sqrt{\tau_+ - s} + \sqrt{\tau_+ - \tau_0}}, \quad \tau_0 = \sqrt{\tau_+} (\sqrt{\tau_+} - \sqrt{\tau_+ - \tau_-}), \quad \tau_{\pm} = (M_B \pm M_K)^2.$$

This parametrization depends on the central value of the form factor at $q^2 = 0$, $f_+(0)$, and the slope parameter b_1^+ (and $M_{\text{res},+} = 5.412$ GeV). At large recoil, the dipole form factor f_T is replaced by the large-energy universal form factor $\xi_P \equiv f_+$ [48, 63]. At low recoil, the dipole form factor f_T is substituted for by means of the improved Isgur-Wise relation [59].

In addition, we vary the decay constant f_{B_s} of the B_s meson, since it constitutes the dominant uncertainty in the decay $B_s \rightarrow \bar{\mu} \mu$. The most recent lattice results [104, 105] have been averaged [106], yielding the number listed in Tab. 12.

In order to assess the dependence of the fit on the choice of priors, we adopt two sets of priors. The first set reflects the uncertainties as reported by the authors of [41, 106, 107],

parameter	central	nominal		wide	
		1σ	support	1σ	support
ζ_{V,A_1,A_2}	1.0	0.15	3σ	0.45	3σ
$f_+(0)$	0.34	[0.32, 0.39]	[0.28, 0.49]	[0.28, 0.49]	[0.0, 0.79]
b_1^+	-2.1	[-3.7, -1.2]	[-6.9, 0.6]	[-6.9, 0.6]	[-10, 3.7]
f_{B_s}	227.7 MeV	6.2 MeV	3σ	18.6 MeV	3σ
$\zeta_{K^*}^{ij}, \zeta_K$	1.0	0.15	3σ	0.45	[0.0, 2.0]
$ r_{0,\perp,\parallel} , r_K $	0.0	0.15	3σ	0.45	3σ

Table 12. Priors of the nuisance parameters of the $B \rightarrow K^{(*)}$ form factors, the B_s decay constant f_{B_s} , and parametrization of lacking subleading corrections at low q^2 ($i = L, R$ and $j = 0, \perp, \parallel$) and high q^2 , specified for the nominal and wide set. All priors are Gaussian and we give the central value, its standard deviation σ , and the support of the prior. The nominal 1σ ranges of V and $A_{1,2}$ correspond to uncertainties quoted in [107], whereas, $f_+(0)$ and b_1^+ are taken from the LCSR analysis [41]; however, possible correlations among both are not available.

thereby assuming the extrapolation of form factors to high q^2 has the same uncertainties as predicted by LCSR's at low q^2 . In the second set we triple the uncertainties. Both sets are given in Tab. 12.

B.3 Subleading Λ/m_b Corrections

There are several distinct sources of Λ/m_b corrections arising in exclusive $B \rightarrow K^{(*)}\bar{\ell}\ell$ decays. Here Λ is assumed to be of the order of the strong scale, however the particular physical meaning depends on the framework. When using power counting we use the generic value of 500 MeV.

The first type is due to the form factor relations in the limit of heavy quark masses [46], which is valid for the whole q^2 -kinematic region. At the leading order in Λ/m_b , they relate the $B \rightarrow K^*$ ($B \rightarrow K$) tensor form factors $T_{1,2,3}$ (f_T) to vector V (f_+) and axial-vector $A_{1,2}$ form factors⁵. This approximation receives a further numerical suppression due to $\mathcal{C}_7/\mathcal{C}_9 \sim \mathcal{O}(0.1)$. The additional large energy limit [47, 48] at low q^2 allows us to eliminate another $B \rightarrow K^*$ form factor, introducing an additional subleading uncertainty not suppressed by $\mathcal{C}_7/\mathcal{C}_9$. Besides subleading corrections due to the use of form factor relations, the two distinct expansions in Λ/m_b , QCDF at low q^2 and the OPE at high q^2 , introduce a second type at the amplitude level, when truncating the expansion after the leading order in Λ/m_b .

At low q^2 , QCDF (or equivalently SCET) provides a possibility to calculate such corrections, which are in general suppressed by a factor of Λ/m_b ⁶. In principle, the partially known corrections [40, 81] could be included as an estimate of the lacking corrections, but

⁵The authors [54] take the viewpoint, that such corrections can be accounted for at low q^2 , if form factor relations are not used in the leading-order contribution (in Λ/m_b and α_s) to the amplitude.

⁶In some subleading corrections one encounters infrared divergences [81, 116].

here we model them by 6 real scale factors for each of the transversity amplitudes $A_{\perp,\parallel,0}^{L,R}$ in the case of $B \rightarrow K^* \bar{\ell} \ell$ and one for $B \rightarrow K \bar{\ell} \ell$. These scale factors $\zeta_{K^*}^{ij}$ ($i = L, R$ and $j = 0, \perp, \parallel$) and ζ_K have Gaussian prior distributions each with central value 1 and a 1σ range of $0.15 \approx \Lambda/m_b$ with a support up to 3σ and include the subleading corrections due to form factor relations discussed above. A 1σ range of $0.45 \approx \Lambda/m_b$ with a support $[0.0, 2.0]$ is chosen for the wide-prior scenario.

At high q^2 , the interaction of the 4-quark operators and the electromagnetic current, which couples to the pair of leptons, might be treated within a local operator product expansion either in full QCD [45] or with subsequent matching on HQET [44]. In both approaches, subleading corrections to the decay amplitudes arise at $(\Lambda/m_b)^2$ and $\alpha_s \Lambda/m_b$, respectively, which are of similar numerical size. The additional suppression factor of Λ/m_b or α_s , yields smaller theory uncertainties due to omission of subleading corrections at high q^2 in contrast to the low- q^2 region. This is also not spoiled by the use of form factor relations [44, 57] for tensor form factors $T_{1,2,3}$ (f_T) due to the accompanying numerical suppression by $\mathcal{C}_7/\mathcal{C}_9$, which depends on the new physics contributions. Note that for both approaches, full QCD and HQET, the subleading corrections are known in part, and in the future it is conceivable that they can be included completely. For example, the unknown subleading form factor arising in [45] could be calculated on the lattice. We follow [44], using $\alpha_s(m_b) \sim 0.3$. This gives rise to 3 complex $r_a \sim \Lambda/m_b$ ($a = 0, \perp, \parallel$) for $B \rightarrow K^* \bar{\ell} \ell$ [59] and one complex $r_K \sim \Lambda/m_b$ for $B \rightarrow K \bar{\ell} \ell$ [64], which are additive at the level of the amplitude. We treat the complex-valued subleading nuisance parameters r_a with eight additional real-valued degrees of freedom, using Gaussian priors each with central value 0, a 1σ range of $0.15 \approx \Lambda/m_b$, and a support up to 3σ for its magnitude. The accompanying phases have uniform priors in $[-\pi/2, \pi/2]$. A three-times-wider 1σ range of $0.45 \approx \Lambda/m_b$ and a support up to 3σ is chosen for the wide-prior scenario.

The choices are also listed in Tab. 12.

C Standard Model Predictions

In this appendix we provide q^2 -integrated SM predictions for measured and unmeasured observables, focusing on those low- and high- q^2 bins that are currently used in experimental analysis and are also accessible to theoretical methods. All quantities are CP averaged and lepton-mass effects have been taken into account using $\ell = \mu$. The theory uncertainties have been obtained using the (nominal) prior distributions of the nuisance parameters. The results are listed in Tab. 13 and Tab. 14 for low and high q^2 . The central value corresponds to the mode and the errors to the smallest 68% interval of the probability distribution obtained with the Monte Carlo method. The value in parentheses is obtained when setting all nuisance parameters to the most probable prior value.

At low q^2 , we do not predict $J_{3,9}$ and associated optimized observables $A_T^{(2,\text{im})}$, since they vanish at leading order in QCDF (including the α_s corrections), although we obtain non-vanishing values due to the implementation of subleading terms of kinematic origin ($\sim M_{K^*}/M_B$).

Observable	[2.0, 4.3]	[1.0, 6.0]
$\langle \mathcal{B}_K \rangle \times 10^7 \text{ }^\dagger$	$0.85^{+0.25}_{-0.13} (0.81)$	$1.85^{+0.54}_{-0.28} (1.75)$
$\langle \mathcal{B}_{K^*} \rangle \times 10^7 \text{ }^\ddagger$	$0.69^{+0.77}_{-0.41} (1.05)$	$1.64^{+1.80}_{-0.83} (2.46)$
$\langle A_{\text{FB}} \rangle$	$0.055^{+0.087}_{-0.033} (0.086)$	$0.03^{+0.07}_{-0.02} (0.05)$
$\langle F_L \rangle$	$0.85^{+0.08}_{-0.20} (0.78)$	$0.81^{+0.09}_{-0.22} (0.73)$
$\langle J_{1s} \rangle \times 10^8$	$1.18^{+0.48}_{-0.35} (1.26)$	$3.43^{+1.37}_{-0.95} (3.66)$
$\langle J_{1c} \rangle \times 10^7$	$0.31^{+0.57}_{-0.29} (0.63)$	$0.83^{+1.07}_{-0.76} (1.37)$
$\langle J_{2s} \rangle \times 10^8$	$0.39^{+0.16}_{-0.12} (0.42)$	$1.13^{+0.45}_{-0.31} (1.21)$
$\langle J_{2c} \rangle \times 10^7$	$-0.30^{+0.28}_{-0.56} (-0.61)$	$-0.79^{+0.75}_{-1.05} (-1.33)$
$\langle J_4 \rangle \times 10^8$	$0.57^{+0.39}_{-0.24} (0.77)$	$1.43^{+0.82}_{-0.62} (1.82)$
$\langle J_5 \rangle \times 10^8$	$-0.69^{+0.37}_{-0.64} (-1.07)$	$-1.80^{+0.88}_{-1.37} (-2.58)$
$\langle J_{6s} \rangle \times 10^8$	$0.84^{+0.45}_{-0.29} (0.90)$	$1.19^{+0.87}_{-0.74} (1.21)$
$\langle J_7 \rangle \times 10^9$	$2.52^{+1.50}_{-1.06} (2.78)$	$5.86^{+3.03}_{-2.62} (6.21)$
$\langle J_8 \rangle \times 10^9$	$-0.89^{+0.49}_{-0.57} (-0.97)$	$-1.79^{+0.94}_{-1.36} (-2.14)$
$\langle A_T^{(3)} \rangle$	$0.45^{+0.12}_{-0.08} (0.50)$	$0.42^{+0.11}_{-0.08} (0.47)$
$\langle A_T^{(4)} \rangle$	$0.63^{+0.17}_{-0.17} (0.69)$	$0.64^{+0.18}_{-0.15} (0.71)$
$\langle A_T^{(5)} \rangle$	$0.41^{+0.03}_{-0.05} (0.42)$	$0.48^{+0.01}_{-0.03} (0.48)$
$\langle A_T^{(\text{re})} \rangle$	$0.61^{+0.10}_{-0.13} (0.54)$	$0.29^{+0.14}_{-0.14} (0.25)$
$\langle H_T^{(1)} \rangle$	$0.45^{+0.08}_{-0.08} (0.48)$	$0.42^{+0.07}_{-0.07} (0.45)$
$\langle H_T^{(2)} \rangle$	$-0.29^{+0.08}_{-0.08} (-0.34)$	$-0.29^{+0.07}_{-0.07} (-0.33)$

Table 13. SM predictions of q^2 -integrated observables at low- q^2 in the bins $q^2 \in [q_{\min}^2, q_{\max}^2]$ for $^\dagger B^- \rightarrow K^- \bar{\mu} \mu$ and $^\ddagger \bar{B}^0 \rightarrow \bar{K}^{*0} \bar{\mu} \mu$. We list the mode and the smallest 68% interval of the probability distribution, along with the value obtained by the conventional method of setting all nuisance parameters to the prior modes (in parentheses).

At high q^2 , $J_{7,8,9}$ is zero at leading order in the OPE and when applying form factor relations, so is $A_T^{(\text{im})}$. Furthermore, we recall that F_L and $A_T^{(2,3)}$ become short-distance independent [57] within the framework of the SM operator basis, and predictions are strongly dependent on the extrapolation of the form factor results from low q^2 obtained using LCSR.

We do not predict J_{6c} since it vanishes in the absence of scalar and tensor operators.

D Distributions

D.1 Amoroso Distribution

Consider the posterior $P(x|D)$, describing the search for a decay whose existence has not been established yet, with x representing the branching ratio. Suppose we know $P(x|D)$

Observable	[14.18, 16.0]	[> 16.0]	[> 14.18]
$\langle \mathcal{B}_K \rangle \times 10^7 \text{ }^\dagger$	$0.39^{+0.22}_{-0.09} (0.37)$	$0.73^{+0.43}_{-0.22} (0.68)$	$1.11^{+0.66}_{-0.28} (1.04)$
$\langle \mathcal{B}_{K^*} \rangle \times 10^7 \text{ }^\ddagger$	$1.19^{+0.37}_{-0.31} (1.26)$	$1.41^{+0.40}_{-0.38} (1.46)$	$2.57^{+0.80}_{-0.68} (2.72)$
$\langle A_{\text{FB}} \rangle$	$-0.44^{+0.07}_{-0.07} (-0.44)$	$-0.37^{+0.06}_{-0.07} (-0.38)$	$-0.40^{+0.06}_{-0.07} (-0.41)$
$\langle F_L \rangle$	$0.38^{+0.04}_{-0.06} (0.36)$	$0.35^{+0.02}_{-0.03} (0.34)$	$0.36^{+0.04}_{-0.05} (0.35)$
$\langle J_{1s} \rangle \times 10^8$	$4.44^{+0.96}_{-1.00} (4.51)$	$5.10^{+1.48}_{-1.11} (5.44)$	$9.70^{+2.31}_{-2.21} (9.96)$
$\langle J_{1c} \rangle \times 10^8$	$3.23^{+1.31}_{-1.37} (3.43)$	$3.40^{+1.41}_{-1.07} (3.72)$	$6.64^{+2.75}_{-2.43} (7.14)$
$\langle J_{2s} \rangle \times 10^8$	$1.48^{+0.32}_{-0.33} (1.50)$	$1.70^{+0.49}_{-0.37} (1.81)$	$3.23^{+0.77}_{-0.74} (3.31)$
$\langle J_{2c} \rangle \times 10^8$	$-3.21^{+1.36}_{-1.31} (-3.41)$	$-3.38^{+1.07}_{-1.41} (-3.70)$	$-6.61^{+2.42}_{-2.74} (-7.11)$
$\langle J_3 \rangle \times 10^8$	$-0.99^{+0.59}_{-0.71} (-1.11)$	$-2.12^{+0.89}_{-0.82} (-2.19)$	$-3.06^{+1.44}_{-1.57} (-3.29)$
$\langle J_4 \rangle \times 10^8$	$2.47^{+0.95}_{-0.85} (2.65)$	$3.10^{+1.08}_{-0.96} (3.27)$	$5.49^{+2.06}_{-1.77} (5.92)$
$\langle J_5 \rangle \times 10^8$	$-3.36^{+0.87}_{-0.87} (-3.54)$	$-2.95^{+0.63}_{-0.80} (-3.17)$	$-6.23^{+1.34}_{-1.79} (-6.72)$
$\langle J_{6s} \rangle \times 10^7$	$-0.52^{+0.10}_{-0.12} (-0.55)$	$-0.53^{+0.11}_{-0.12} (-0.56)$	$-1.05^{+0.22}_{-0.24} (-1.11)$
$\langle A_T^{(2)} \rangle$	$-0.38^{+0.17}_{-0.18} (-0.37)$	$-0.64^{+0.15}_{-0.10} (-0.60)$	$-0.51^{+0.16}_{-0.16} (-0.50)$
$\langle A_T^{(3)} \rangle$	$1.45^{+0.29}_{-0.31} (1.47)$	$1.95^{+0.42}_{-0.40} (2.01)$	$1.67^{+0.36}_{-0.34} (1.72)$
$\langle A_T^{(4)} \rangle$	$0.66^{+0.14}_{-0.14} (0.67)$	$0.48^{+0.10}_{-0.10} (0.48)$	$0.56^{+0.12}_{-0.11} (0.57)$
$\langle A_T^{(5)} \rangle$	$0.085^{+0.008}_{-0.008} (0.081)$	$0.111^{+0.014}_{-0.014} (0.109)$	$0.123^{+0.012}_{-0.012} (0.120)$
$\langle A_T^{(\text{re})} \rangle$	$-0.982^{+0.110}_{-0.003} (-0.915)$	$-0.777^{+0.099}_{-0.089} (-0.767)$	$-0.843^{+0.075}_{-0.087} (-0.834)$
$\langle H_T^{(1)} \rangle$	$0.9996^{+0.0002}_{-0.0003} (0.9996)$	$0.9986^{+0.0008}_{-0.0007} (0.9986)$	$0.9970^{+0.0017}_{-0.0018} (0.9969)$
$\langle H_T^{(2)} \rangle$	$-0.9844^{+0.0027}_{-0.0020} (-0.9853)$	$-0.9719^{+0.0034}_{-0.0024} (-0.9722)$	$-0.9748^{+0.0040}_{-0.0031} (-0.9751)$
$\langle H_T^{(3)} \rangle$	$-0.9837^{+0.0024}_{-0.0018} (-0.9845)$	$-0.9614^{+0.0017}_{-0.0011} (-0.9618)$	$-0.9606^{+0.0018}_{-0.0016} (-0.9613)$

Table 14. SM predictions of q^2 -integrated observables at high- q^2 in the bins $q^2 \in [q_{\min}^2, q_{\max}^2]$ for $^\dagger B^- \rightarrow K^- \bar{\mu} \mu$ and $^\ddagger \bar{B}^0 \rightarrow \bar{K}^{*0} \bar{\mu} \mu$. We list the mode and the smallest 68% interval of the probability distribution, along with the value obtained by the conventional method of setting all nuisance parameters to the prior modes (in parentheses).

at a number of data points, (x_i, P_i) . Using the cumulative distribution function

$$F(x_a|D) = \int_{-\infty}^{x_a} dx P(x|D), \quad (\text{D.1})$$

we can determine the limit x_a at level a from $F(x_a|D) = a$. For convenience, we seek an analytical expression $g(x)$ interpolating the data points. We constrain $g(\cdot)$ by requiring that it vanish for negative branching ratios and that it yield the same 10[50,90]% limits as obtained from $F(\cdot|D)$:

$$g(x \leq 0) = 0 \quad (\text{D.2})$$

$$\int_0^{x_a} dx g(x) = a, \quad a = 0.1, 0.5, 0.9. \quad (\text{D.3})$$

We choose $g(x) = \text{Amoroso}(x|l, \lambda, \alpha, \beta)$. The Amoroso family [99] is a continuous unimodal four-parameter family of probability distributions that easily accommodates the

constraints and provides an accurate approximation. Many well known distributions are direct members or appear as limits of the Amoroso family. Its functional form is

$$\text{Amoroso}(x|l, \lambda, \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \left| \frac{\beta}{\lambda} \right| \left(\frac{x-l}{\lambda} \right)^{\alpha\beta-1} \exp \left[- \left(\frac{x-l}{\lambda} \right)^\beta \right] \quad (\text{D.4})$$

for $x, l, \lambda, \alpha, \beta \in \mathbb{R}, \alpha > 0$,
support $x \geq l$ if $\lambda > 0$, $x \leq l$ if $\lambda < 0$.

We set the location parameter l to the minimum physical value, $l = 0$, and ensure that $\lambda > 0$ to satisfy (D.2). The scale parameter λ and the shape parameters α and β are found by numerically solving the set of three equations (D.3). In the limit of $\alpha \rightarrow \infty$ and $\beta = 1$, Amoroso(\cdot) converges to a Gaussian distribution [99].

D.2 LogGamma Distribution

Consider a nuisance parameter ν whose reported uncertainties are asymmetric, $\nu = \mu_{-a}^{+b}, a \neq b$. In this case, we use the LogGamma distribution [99] to obtain a continuous prior over the given range of ν . The LogGamma family is a continuous unimodal three-parameter family of probability distributions

$$\text{LogGamma}(\nu|l, \lambda, \alpha) = \frac{1}{\Gamma(\alpha)|\lambda|} \exp \left(\alpha \left(\frac{\nu-l}{\lambda} \right) - \exp \left(\frac{\nu-l}{\lambda} \right) \right) \quad (\text{D.5})$$

for $\nu, l, \lambda, \alpha \in \mathbb{R}, \alpha > 0$,
support $-\infty \leq \nu \leq \infty$.

The three parameters are uniquely fixed by demanding that the mode of $P(\nu)$ be at μ , that the interval $[\mu - a, \mu + b]$ contain 68%, and that the density be identical at $\mu - a$ and $\mu + b$. More concisely, we have three conditions:

$$\arg \max_{\nu} P(\nu) = \mu \quad (\text{D.6})$$

$$\int_{\mu-a}^{\mu+b} d\nu P(\nu) = 0.68 \quad (\text{D.7})$$

$$P(\mu - a) = P(\mu + b). \quad (\text{D.8})$$

While the first constraint is used to fix the location parameter l , the scale parameter λ and the shape parameter α must be extracted numerically by solving the coupled equations (D.7) and (D.8). For a finite range of ν , say $[\nu_{min}, \nu_{max}]$, the resulting density is normalized such that $\int_{\nu_{min}}^{\nu_{max}} d\nu P(\nu) = 1$.

The asymmetry is governed by α : LogGamma(\cdot) approaches a Gaussian distribution in the limit $\alpha \rightarrow \infty$.

References

- [1] T. E. Coan *et al.* [CLEO Collaboration], Phys. Rev. Lett. **84** (2000) 5283 [hep-ex/9912057].
- [2] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. Lett. **93** (2004) 081802 [arXiv:hep-ex/0404006].

- [3] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. D **73** (2006) 092001 [arXiv:hep-ex/0604007].
- [4] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. D **79** (2009) 031102 [arXiv:0804.4412 [hep-ex]].
- [5] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. D **78** (2008) 071102 [arXiv:0807.3103 [hep-ex]].
- [6] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. Lett. **102** (2009) 091803 [arXiv:0807.4119 [hep-ex]].
- [7] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. Lett. **103** (2009) 211802 [arXiv:0906.2177 [hep-ex]].
- [8] [BABAR Collaboration], arXiv:1204.3933 [hep-ex].
- [9] V. Poireau [BaBar Collaboration], arXiv:1205.2201 [hep-ex].
- [10] M. Nakao *et al.* [BELLE Collaboration], Phys. Rev. D **69** (2004) 112001 [hep-ex/0402042].
- [11] M. Iwasaki *et al.* [Belle Collaboration], Phys. Rev. D **72** (2005) 092005 [arXiv:hep-ex/0503044].
- [12] Y. Ushiroda *et al.* [Belle Collaboration], Phys. Rev. D **74** (2006) 111104 [hep-ex/0608017].
- [13] J. T. Wei *et al.* [BELLE Collaboration], Phys. Rev. Lett. **103** (2009) 171801 [arXiv:0904.0770 [hep-ex]].
- [14] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. Lett. **106** (2011) 161801 [arXiv:1101.1028 [hep-ex]].
- [15] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. Lett. **107** (2011) 201802 [arXiv:1107.3753 [hep-ex]].
- [16] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. Lett. **108** (2012) 081807 [arXiv:1108.0695 [hep-ex]].
- [17] R. Aaij *et al.* [LHCb Collaboration], arXiv:1112.3515 [hep-ex].
- [18] C. Parkinson [LHCb Collaboration], LHCb-CONF-2012-008, Presented at 47th Rencontres de Moriond 2012: QCD Sessions.
- [19] V. M. Abazov *et al.* [D0 Collaboration], Phys. Lett. B **693** (2010) 539 [arXiv:1006.3469 [hep-ex]].
- [20] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. Lett. **107** (2011) 239903 [Phys. Rev. Lett. **107** (2011) 191801] [arXiv:1107.2304 [hep-ex]].
- [21] H. Miyake [CDF Collaboration], Presented at La Thuile, Aosta Valley, Italy, February 26 - March 3, 2012.
- [22] R. Aaij *et al.* [LHCb Collaboration], Phys. Lett. B **708** (2012) 55 [arXiv:1112.1600 [hep-ex]].
- [23] R. Aaij *et al.* [LHCb Collaboration], arXiv:1203.4493 [hep-ex].
- [24] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Rev. Lett. **107** (2011) 191802 [arXiv:1107.5834 [hep-ex]].
- [25] S. Chatrchyan *et al.* [CMS Collaboration], arXiv:1203.3976 [hep-ex].
- [26] G. Aad *et al.* [The ATLAS Collaboration], arXiv:1204.0735 [hep-ex].

- [27] B. Adeva *et al.* [The LHCb Collaboration], arXiv:0912.4179 [hep-ex].
- [28] K. G. Chetyrkin, M. Misiak and M. Munz, Phys. Lett. B **400** (1997) 206 [Erratum-ibid. B **425** (1998) 414] [arXiv:hep-ph/9612313].
- [29] M. Misiak *et al.*, Phys. Rev. Lett. **98** (2007) 022002 [arXiv:hep-ph/0609232].
- [30] M. Misiak and M. Steinhauser, Nucl. Phys. B **764** (2007) 62 [arXiv:hep-ph/0609241].
- [31] C. Bobeth, P. Gambino, M. Gorbahn and U. Haisch, JHEP **0404** (2004) 071 [arXiv:hep-ph/0312090].
- [32] T. Huber, E. Lunghi, M. Misiak and D. Wyler, Nucl. Phys. B **740** (2006) 105 [arXiv:hep-ph/0512066].
- [33] T. Huber, T. Hurth and E. Lunghi, Nucl. Phys. B **802** (2008) 40 [arXiv:0712.3009 [hep-ph]].
- [34] C. Greub, V. Pilipp and C. Schupbach, JHEP **0812** (2008) 040 [arXiv:0810.4077 [hep-ph]].
- [35] T. Aushev, W. Bartel, A. Bondar, J. Brodzicka, T. E. Browder, P. Chang, Y. Chao and K. F. Chen *et al.*, arXiv:1002.5012 [hep-ex].
- [36] B. O’Leary *et al.* [SuperB Collaboration], arXiv:1008.1541 [hep-ex].
- [37] C. Bobeth, G. Hiller and D. van Dyk, J. Phys. Conf. Ser. **335** (2011) 012038 [arXiv:1105.2659 [hep-ph]].
- [38] M. Beneke, T. Feldmann and D. Seidel, Nucl. Phys. B **612** (2001) 25 [arXiv:hep-ph/0106067].
- [39] S. W. Bosch and G. Buchalla, Nucl. Phys. B **621** (2002) 459 [hep-ph/0106081].
- [40] M. Beneke, T. Feldmann and D. Seidel, Eur. Phys. J. C **41** (2005) 173 [arXiv:hep-ph/0412400].
- [41] A. Khodjamirian, T. Mannel, A. A. Pivovarov and Y. -M. Wang, JHEP **1009** (2010) 089 [arXiv:1006.4945 [hep-ph]].
- [42] A. Ali, G. Kramer and G. h. Zhu, Eur. Phys. J. C **47** (2006) 625 [arXiv:hep-ph/0601034].
- [43] K. S. M. Lee, Z. Ligeti, I. W. Stewart and F. J. Tackmann, Phys. Rev. D **75** (2007) 034016 [arXiv:hep-ph/0612156].
- [44] B. Grinstein and D. Pirjol, Phys. Rev. D **70** (2004) 114005 [arXiv:hep-ph/0404250].
- [45] M. Beylich, G. Buchalla and T. Feldmann, Eur. Phys. J. C **71** (2011) 1635 [arXiv:1101.5118 [hep-ph]].
- [46] N. Isgur and M. B. Wise, Phys. Rev. D **42** (1990) 2388.
- [47] J. Charles, A. Le Yaouanc, L. Oliver, O. Pene and J. C. Raynal, Phys. Rev. D **60** (1999) 014001 [hep-ph/9812358].
- [48] M. Beneke and T. Feldmann, Nucl. Phys. B **592** (2001) 3 [arXiv:hep-ph/0008255].
- [49] B. Grinstein and D. Pirjol, Phys. Lett. B **533** (2002) 8 [arXiv:hep-ph/0201298].
- [50] F. Kruger, L. M. Sehgal, N. Sinha and R. Sinha, Phys. Rev. D **61** (2000) 114028 [Erratum-ibid. D **63** (2001) 019901] [hep-ph/9907386].
- [51] F. Kruger and J. Matias, Phys. Rev. D **71** (2005) 094009 [hep-ph/0502060].
- [52] C. Bobeth, G. Hiller and G. Piranishvili, JHEP **0807** (2008) 106 [arXiv:0805.2525 [hep-ph]].
- [53] U. Egede, T. Hurth, J. Matias, M. Ramon and W. Reece, JHEP **0811** (2008) 032 [arXiv:0807.2589 [hep-ph]].

- [54] W. Altmannshofer, P. Ball, A. Bharucha, A. J. Buras, D. M. Straub and M. Wick, JHEP **0901** (2009) 019 [arXiv:0811.1214 [hep-ph]].
- [55] A. Bharucha and W. Reece, Eur. Phys. J. C **69** (2010) 623 [arXiv:1002.4310 [hep-ph]].
- [56] U. Egede, T. Hurth, J. Matias, M. Ramon and W. Reece, JHEP **1010** (2010) 056 [arXiv:1005.0571 [hep-ph]].
- [57] C. Bobeth, G. Hiller and D. van Dyk, JHEP **1007** (2010) 098 [arXiv:1006.5013 [hep-ph]].
- [58] A. K. Alok, A. Datta, A. Dighe, M. Duraisamy, D. Ghosh and D. London, JHEP **1111** (2011) 121 [arXiv:1008.2367 [hep-ph]].
- [59] C. Bobeth, G. Hiller and D. van Dyk, JHEP **1107** (2011) 067 [arXiv:1105.0376 [hep-ph]].
- [60] D. Becirevic and E. Schneider, Nucl. Phys. B **854** (2012) 321 [arXiv:1106.3283 [hep-ph]].
- [61] J. Matias, F. Mescia, M. Ramon and J. Virto, arXiv:1202.4266 [hep-ph].
- [62] C. Bobeth, T. Ewerth, F. Kruger and J. Urban, Phys. Rev. D **64** (2001) 074014 [hep-ph/0104284].
- [63] C. Bobeth, G. Hiller and G. Piranishvili, JHEP **0712** (2007) 040 [arXiv:0709.4174 [hep-ph]].
- [64] C. Bobeth, G. Hiller, D. van Dyk and C. Wacker, JHEP **1201** (2012) 107 [arXiv:1111.2558 [hep-ph]].
- [65] S. Descotes-Genon, D. Ghosh, J. Matias, M. Ramon, JHEP **1106** (2011) 099 [arXiv:1104.3342 [hep-ph]].
- [66] W. Altmannshofer, P. Paradisi and D. M. Straub, JHEP **1204** (2012) 008 [arXiv:1111.1257 [hep-ph]].
- [67] D. van Dyk *et al.*, [EOS Collaboration],
<http://project.het.physik.tu-dortmund.de/eos/>. The particular source code used for this work can be obtained from the authors on demand.
- [68] C. Bobeth, M. Misiak and J. Urban, Nucl. Phys. B **574** (2000) 291 [arXiv:hep-ph/9910220].
- [69] M. Misiak and M. Steinhauser, Nucl. Phys. B **683** (2004) 277 [arXiv:hep-ph/0401041].
- [70] M. Gorbahn and U. Haisch, Nucl. Phys. B **713** (2005) 291 [arXiv:hep-ph/0411071].
- [71] M. Gorbahn, U. Haisch and M. Misiak, Phys. Rev. Lett. **95** (2005) 102004 [arXiv:hep-ph/0504194].
- [72] M. Czakon, U. Haisch and M. Misiak, JHEP **0703** (2007) 008 [arXiv:hep-ph/0612329].
- [73] A. J. Buras, P. Gambino and U. A. Haisch, Nucl. Phys. B **570** (2000) 117 [arXiv:hep-ph/9911250].
- [74] P. Gambino and U. Haisch, JHEP **0110** (2001) 020 [arXiv:hep-ph/0109058].
- [75] F. Borzumati, C. Greub, T. Hurth and D. Wyler, Phys. Rev. D **62** (2000) 075005 [hep-ph/9911245].
- [76] G. Hiller and F. Kruger, Phys. Rev. D **69** (2004) 074020 [hep-ph/0310219].
- [77] C. Bobeth and U. Haisch, arXiv:1109.1826 [hep-ph].
- [78] F. Muheim, Y. Xie and R. Zwicky, Phys. Lett. B **664** (2008) 174 [arXiv:0802.0876 [hep-ph]].
- [79] E. Kou, A. Le Yaouanc and A. Tayduganov, Phys. Rev. D **83** (2011) 094007 [arXiv:1011.6593 [hep-ph]].

- [80] A. Tayduganov, E. Kou and A. Le Yaouanc, Phys. Rev. D **85** (2012) 074011 [arXiv:1111.6307 [hep-ph]].
- [81] T. Feldmann and J. Matias, JHEP **0301** (2003) 074 [arXiv:hep-ph/0212158].
- [82] C. Hambroek and G. Hiller, arXiv:1204.4444 [hep-ph].
- [83] K. de Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino and N. Tuning, arXiv:1204.1737 [hep-ph].
- [84] R. Aaij *et al.* [LHCb Collaboration], LHCb-CONF-2012-002.
- [85] R. Aaij *et al.* [LHCb Collaboration], Phys. Rev. Lett. **108** (2012) 241801 [arXiv:1202.4717 [hep-ex]].
- [86] F. Beaujean, PhD thesis (2012), to appear.
- [87] A. Gelman and D. B. Rubin, Statist. Sci. **7** (1992) 457.
- [88] J. Goldberger and S. Roweis, Advances in Neural Information Processing Systems **17** (2004) 505.
- [89] O. Cappé, R. Douc, A. Guillin, J. M. Marin and C. P. Robert, Statistics and Computing **18** (2008) 447 [arXiv:0710.4242 [stat.CO]].
- [90] D. Wraith, M. Kilbinger, K. Benabed, O. Cappé, J. -F. Cardoso, G. Fort, S. Prunet and C. P. Robert, Phys. Rev. D **80** (2009) 023507 [arXiv:0903.0837 [astro-ph.CO]].
- [91] K. Nakamura *et al.* [Particle Data Group], J. Phys. G **37** (2010) 075021.
- [92] M. Bona *et al.* [UTfit Collaboration], JHEP **0610** (2006) 081 [hep-ph/0606167]. We use the data from Summer 2010, <http://utfit.org>.
- [93] E. T. Jaynes and G. L. Bretthorst, *Probability theory*, Cambridge University Press (2003).
- [94] A. L. Read, J. Phys. G **28** (2002) 2693.
- [95] R. R. de Austri, R. Trotta and L. Roszkowski, JHEP **0605** (2006) 002 [hep-ph/0602028].
- [96] H. Flacher, M. Goebel, J. Haller, A. Hocker, K. Monig and J. Stelzer, Eur. Phys. J. C **60** (2009) 543 [Erratum-ibid. C **71** (2011) 1718] [arXiv:0811.0009 [hep-ph]].
- [97] W. R. Reece, CERN-THESIS-2010-095.
- [98] J. Heinrich, CDF-MEMO **7587** (2005).
- [99] G. Crooks, arXiv:1005.3274 [math-st].
- [100] F. Beaujean, A. Caldwell, D. Kollar and K. Kröninger, Phys. Rev. D **83** (2011) 012004.
- [101] C. Bobeth, G. Hiller and D. van Dyk, to appear.
- [102] [Tevatron Electroweak Working Group and CDF Collaboration and D0 Collab], arXiv:0903.2503 [hep-ex].
- [103] J. Simone *et al.* [Fermilab Lattice and MILC Collaboration], PoS LATTICE **2010** (2010) 317.
- [104] C. McNeile, C. T. H. Davies, E. Follana, K. Hornbostel and G. P. Lepage, Phys. Rev. D **85** (2012) 031503 [arXiv:1110.4510 [hep-lat]].
- [105] A. Bazavov *et al.* [Fermilab Lattice and MILC Collaboration], arXiv:1112.3051 [hep-lat].

- [106] J. Laiho, E. Lunghi and R. S. Van de Water, Phys. Rev. D **81** (2010) 034503 [arXiv:0910.2928 [hep-ph]]. We use the most recent update available from <http://krone.physik.unizh.ch/~lunghi/webpage/LatAves/page7/page7.html#bottomsemileptonic>
- [107] P. Ball and R. Zwicky, Phys. Rev. D **71** (2005) 014029 [arXiv:hep-ph/0412079].
- [108] J. Charles *et al.* [CKMfitter Group Collaboration], Eur. Phys. J. C **41** (2005) 1 [hep-ph/0406184]. Updated results and plots available at: <http://ckmfitter.in2p3.fr>.
- [109] D. Becirevic, V. Lubicz and F. Mescia, Nucl. Phys. B **769** (2007) 31 [hep-ph/0611295].
- [110] A. Al-Haydari *et al.* [QCDSF Collaboration], Eur. Phys. J. A **43** (2010) 107 [arXiv:0903.1664 [hep-lat]].
- [111] Z. Liu, S. Meinel, A. Hart, R. R. Horgan, E. H. Muller and M. Wingate, PoS LAT **2009** (2009) 242 [arXiv:0911.2370 [hep-lat]].
- [112] R. Zhou *et al.* [Fermilab Lattice and MILC Collaborations], arXiv:1111.0981 [hep-lat].
- [113] Z. Liu, S. Meinel, A. Hart, R. R. Horgan, E. H. Muller and M. Wingate, arXiv:1101.2726 [hep-ph].
- [114] A. Bharucha, T. Feldmann and M. Wick, JHEP **1009** (2010) 090 [arXiv:1004.3249 [hep-ph]].
- [115] C. Bourrely, I. Caprini and L. Lellouch, Phys. Rev. D **79** (2009) 013008 [Erratum-ibid. D **82** (2010) 099902] [arXiv:0807.2722 [hep-ph]].
- [116] A. L. Kagan and M. Neubert, Phys. Lett. B **539** (2002) 227 [hep-ph/0110078].